Concepts and Strategies to Actively Monitor the Risks of Medications in Pregnancy: Enhancing Post-Marketing Surveillance

Workshop Summary

November 29 - 30, 2000
Holiday Inn, Bethesda, Maryland

The workshop was planned in conjunction with representatives from

Division of Birth Defects and Developmental Disabilities,
Centers for Disease Control and Prevention

Division of HIV/AIDS Prevention,
Centers for Disease Control and Prevention

Office of Women's Health,
Centers for Disease Control and Prevention

Office on Women's Health,
Department of Health and Human Services

Center for Drug Evaluation Research,
Food and Drug Administration

Center for Biologics Evaluation Research,
Food and Drug Administration

Office of Women's Health,
Food and Drug Administration

GlaxoSmithKline

International Society of Pharmacoepidemiology

March of Dimes

Office of Research on Women's Health,
National Institutes of Health

Organization of Teratology Information Services

Table of Contents

Background, Purpose, and Format of the Workshop

Since the teratogenic effects of thalidomide were recognized in the 1960s, experimental testing of animals to evaluate the reproductive effects of medications are routinely conducted as a condition of drug approval by the Food and Drug Administration (FDA). These regulatory procedures have helped prevent the introduction of new teratogens in the United States and to relieve widespread concerns that pregnant women will be exposed to a large number of harmful drugs. However, because pregnant women are excluded from clinical trials of drug efficacy, there is little experience with the reproductive effects of most prescription drugs in humans at the time they are marketed. Even when there has been strong indication of potential teratogenicity from animal studies and a drug was introduced anyway, confirmation of the teratogenic effect usually has first been documented by astute clinicians after rather widespread use. This is what happened with Accutane®.

Historically in the United States, the predominant method of collecting data about drug safety in pregnancy has been through spontaneous reporting of adverse outcomes. Clinicians, pharmacists, and patients report adverse events to pharmaceutical and medical product companies, which are required by law to report them to the FDA. While helpful in monitoring for unique and unexpected adverse events, these reports can be biased toward more severe outcomes and do not always accurately reflect the rates of reproductive events in the population. The system is insufficient for understanding the full scope of the effects of medications in pregnancy.

Because there is no comprehensive systematic mechanism in place for evaluating the risks of medications in pregnancy, many pregnant women are advised to avoid all drugs as potentially harmful. While this is generally prudent, a number of maternal conditions require ongoing treatment, the cessation of which could pose a threat to the health of both mother and child. Maternal epilepsy, diabetes, autoimmune disorders, and some psychiatric conditions are examples. In addition, at least half of pregnancies in the United States each year are unplanned so that an unknown number of women inadvertently use prescription and/or non-prescription drugs early in gestation before realizing they are pregnant. These early weeks form the critical period for organ and nervous system development, during which time the embryo can be most vulnerable. Lack of information about the risks and safety of medications can result in unwarranted fear and anxiety and, in the extreme, could lead to termination of a wanted pregnancy. Additional and better information is needed to assist health care providers and pregnant women in making decisions about the management of exposed pregnancies.

In an attempt to respond to these issues, pharmaceutical companies have established pregnancy registries to capture information about the effects of certain prescription drugs. Exposed pregnancies are followed prospectively to determine their outcome and to calculate the frequency of occurrence of physical defects. The first of these was the Acyclovir Pregnancy Registry, which was established in 1984 by what was then BurroughsWellcome (now GlaxoSmithKline) and which employed an advisory committee of specialists from the fields of pediatrics, obstetrics, genetics, teratology, epidemiology, and public health to ensure scientific integrity and objectivity. That registry closed in 1999, but its information was subsequently incorporated into the product label for Acyclovir and is one component cited by clinicians and patients as being the most helpful and informative.

The success of that registry led to the establishment of similar registries for other products commonly used by women of childbearing age, particularly when there was concern about potential effects on the developing fetus. Currently active pregnancy registries exist for a variety of medications, including antiretroviral drugs, antiepileptic drugs, asthma drugs, lamotrigine, montelukast, sumatriptan, naratriptan, rizatriptan, rofecoxib, bupropion, leflunomide, trazodone, nefazodone, imiquimod, and varicella vaccine. FDA has prepared a guidance document for the pharmaceutical industry on how to approach the conduct of pregnancy registries. However, it has become clear that establishing a unique registry for each medication of concern is not necessarily efficient or practical.

In January 1999, the Division of Birth Defects and Developmental Disabilities, Centers for Disease Control (CDC), and the Center for Drug Evaluation and Research, FDA, began a series of discussions on (1) how best to advise pharmaceutical companies about the conduct of pregnancy registries, and (2) on the future of these activities and of post-marketing surveillance for the effects of medication use during pregnancy in general. Mutual areas of concern included the quality of the data being generated, how best to interpret the findings, the burden to clinicians asked to participate in multiple independent registries, and the protection of human subjects. The mutual interest of these agencies in facilitating data collection systems that would ultimately result in information helpful to clinicians and patients, without overloading individuals, companies or organizations, led to a workshop entitled "Concepts and Strategies to Actively Monitor the Effects of Medications in Pregnancy: Enhancing Post-Marketing Surveillance" held November 29-30, 2000.

The workshop brought together 100 participants with a wide variety of backgrounds who have a stake in, and are dedicated to, the safety of medication use during pregnancy. These included health care providers in the specialties of pediatrics, obstetrics, genetics, teratology, genetic counseling and child development; public health specialists; private sector researchers; members of professional organizations, pharmaceutical companies, and government agencies; and consumers. The goals were to (1) explore the issues involved in actively monitoring the effects of medication use during pregnancy; (2) obtain input on ways to improve post-marketing surveillance for the effects of medication use during pregnancy; (3) exchange ideas, tap into creative thinking, and educate each other. A consensus of opinions was not sought and was not obtained.

Five separate discussion groups of 20 people each considered topics in depth and in parallel, then reported back to the larger group. Scheduled presentations were made only to provide background and set the stage. The discussions were divided into three sessions.

1. The first session addressed which medications, outcomes, and levels of risk it is important to monitor and which it is realistic to monitor through post-marketing surveillance.
2. In the second session, three separate models for conducting post-marketing surveillance were proposed, each designed to illustrate specific methodologic approaches and data sources. The groups were asked to discuss the characteristics, strengths, and limitations of each.
3. The third session addressed what other approaches would yield improvements in post-marketing surveillance, how these approaches could be combined or coordinated to ensure that the goals of post-marketing surveillance are achieved, and what the next steps would be.

This document summarizes the major points of the scheduled presentations and the group discussions that took place during the 2-day workshop. It is intended to present the opinions and advice offered by those who attended, and to promote an understanding of the complex issues surrounding surveillance for the effects of medication use during pregnancy.

Presentations

Why We Need To Improve Post-Marketing Surveillance

Christina Chambers, MPH
California Teratogen Information Service and Clinical Research Program

I am going to review the current state of the art of post-marketing surveillance for drug safety during pregnancy in the United States and where the challenges lie. It is important to keep in mind the incredible frequency with which prescription medications are given to pregnant women. A recent study published in *Lancet* and conducted with approximately 1,000 pregnant women in France reported that the average number of prescriptions written for women in the first trimester was 5.1. Over 60% of these were for analgesics, over 50% for cardiovascular medications and antibiotics, and over 20% were for rheumatic and central nervous system medications.[1] Importantly, more than 90% of the prescriptions were for drugs that fall into Food and Drug Administration (FDA) categories indicating a lack of sufficient human data to suggest whether or not use of these drugs is safe during pregnancy.

I'm going to take you through the chronology of information that becomes available on a drug before it is marketed, the typical clinical experiences that occur once a drug is available, and what sources of information are available after a drug is marketed. Prior to a drug's approval, reproductive toxicity studies are conducted in various animal species as a necessary first step in screening new pharmaceuticals. Currently, there are a number of efforts to improve the predictive value of animal studies, but it remains difficult to extrapolate from animal data to human clinical applications. This is true primarily because of differences in susceptibility and sensitivity between animal species and humans and because the dose and route of administration given to animals often is not comparable to that typically used in humans.

The next source of information is from pregnancies that occur during premarketing clinical trials. Although pregnant women are typically excluded from these studies for ethical reasons, limited information can be gained from unintended pregnancies that occur. However, interpretation of the results is limited. An example is the sumatriptan clinical trial, which was actually conducted in the post-marketing arena but which demonstrates some of the strengths and weaknesses of the information obtained from this approach.[2] The trial was to evaluate the effectiveness of injectable sumatriptan among people with migraine headaches, of whom about 10,000 were women. It was anticipated that unintended pregnancies would occur and so a secondary goal of the study was to look at pregnancy outcomes. A total of 168 women became pregnant during the trial; 76 used sumatriptan only after conception and 92 used it prior to conception. Strengths of the study included (1) the precise measurement of exposure timing relative to the last menstrual period; (2) the availability of a disease-matched control group in the form of pregnant women with migraine who used the medication only before conception; and (3) the prospective enrollment of all study participants before they were pregnant, enabling evaluation of early outcomes such as spontaneous abortion. Weaknesses of the study included (1) the fact that women who enroll in and become pregnant during a clinical trial might not be typical of all pregnant women who would use the drug, (2) the lack of a comprehensive standardized method of evaluating pregnancy outcomes, and (3) an inadequate sample size based on incidental pregnancies to provide adequate power for detecting rare outcomes such as major malformations.

Once a drug is on the market, there are several realities that create the situation in which medication exposure during pregnancy is likely to occur. First is the fact that an estimated 56% of pregnancies in the United States each year are unplanned. Second, survey data indicate that fewer than half of all pregnancies are recognized before the fifth week of gestation, and approximately 20% remain unrecognized at up to 8 weeks, well into the time when a significant proportion of embryonic development has been completed.[3, 4] This sets up a common clinical dilemma for a pregnant woman, her partner, and her health care provider. Consider a woman taking lovastatin, a medication for lowering cholesterol, who discovers she is pregnant at 4 weeks ges-

tation. Her obstetrician tells her that there are no controlled human studies about use of this drug during pregnancy, there are several case reports in the literature of congenital anomalies with its use, the animal studies are concerning, the drug is in FDA pregnancy category X, and she does not really need the medication during pregnancy. So, she stops taking it, but then asks, "Am I at increased risk of having a baby with a birth defect or any other sort of problem?" And the answer is, "We don't know."

Another reality is that chronic medical conditions are relatively common among women of reproductive age, and many require continued treatment through part or all of gestation. For example, major clinical depression is thought to occur in maybe 10% of women in this age range, seizure disorders in 0.5%, autoimmune diseases in maybe 1% to 2%, and chronic asthma in 2% to 5%. This sets up another common clinical dilemma. Consider a woman with severe asthma who is well-controlled on a drug regimen that includes montelukast, a relatively new leukotriene receptor antagonist. She is planning a pregnancy and would like to continue using this medication throughout gestation. Her obstetrician tells her that there are no controlled human studies about the use of this drug in pregnancy. But she and her allergist are concerned that her asthma might not be well-controlled if she switches medications and that, if her symptoms worsen, it actually might be detrimental to her baby. So she asks, "Can I feel comfortable continuing to take this drug after I become pregnant?" And the answer is, "We don't know."

Once a drug has been marketed and we know these clinical dilemmas are going to arise, there are several sources of information presently in place. First of these are the programs implemented and administered by the FDA. The Adverse Event Reporting System (AERS) receives reports from drug manufacturers, packers and distributors of all serious adverse events reported to them by health care providers and consumers or published in the literature. FDA regulations mandate these reports be submitted within 15 calendar days of receipt. Major congenital anomalies are considered serious adverse events. There are also periodic reporting requirements for any post-marketing safety studies sponsored by manufacturers. The FDA also administers the MedWatch program, an educational program that promotes reporting of adverse events by health care providers through use of a standard format and mechanism. In addition, consumers and health care providers can report adverse events directly to the FDA.

Strengths of the mandated FDA systems are that (1) they require timely reporting, (2) they draw on a variety of reporters, and (3) major congenital anomalies associated with medication use are reviewed on a case-by-case basis. There is the potential for the FDA to recognize signals that can generate hypotheses for testing with other study methods. One limitation of the FDA systems is that there is little information on the actual denominator involved so that, if 10 neural tube defects are reported with a certain medication, it isn't known how many additional women took the drug and did not have a child with a neural tube defect. This makes it difficult to determine when a signal is truly a signal. In addition, reporting by prescribers themselves is not mandated but relies on the spontaneous initiative of the prescriber or consumer, which can introduce bias. Also, the wide range of capability of reporters leads to varying quality of the reports generated.

Another source of information after a drug is marketed is the industry pregnancy registries. These can be retrospective or prospective in design; can involve a single drug or a class of drugs, as does the Antiretroviral Pregnancy Registry; and can be supported by a single manufacturer or a consortium of manufacturers, as is the North American Antiepileptic Drug Pregnancy Registry. The Fluoxetine Pregnancy Registry is an example of a typical design. This registry was set up by Eli Lilly as a worldwide repository for retrospective and prospective reports on pregnancy exposures. In the 8 years after the drug was marketed, the registry accumulated data on over 2,000 fluoxetine-exposed pregnancies. It is a very commonly used drug. Strengths of this registry include the utilization of standardized methods for data collection on pregnancy exposure, history and outcome, and the accumulation of reports from both physicians and patients in one central location. This provides the manufacturer with the opportunity to look for patterns of defects or other adverse outcomes from which to generate hypotheses for further testing

with other methods. One limitation of the Fluoxetine Pregnancy Registry, and of the registry approach in general, is the lack of an internally generated control group. Pregnancy registries typically use external reference groups, population data, or historical controls that might or might not have characteristics similar to those of women who actually used the drug. Another primary difficulty is the lack of outcome validation. Obstetricians or family physicians typically report pregnancy outcomes to the registries, but are not necessarily the best to convey whether a child had an adverse outcome and what that specific outcome was. In addition, many registries have a very high rate of loss to followup, making it difficult to obtain outcome information on a high proportion of the ascertained pregnancies. Again, these registries rely on spontaneous reporting which can lead to bias.

Another source of post-marketing data comes from the Organization of Teratology Information Services (OTIS). This network of services in North America provides telephone risk counseling to pregnant women and health care providers about all types of exposures, including those to prescription medications. Over the years, some OTIS members, primarily the Mother Risk Program in Toronto and the California group in the United States, have conducted cohort studies on specific medications used during pregnancy. More recently, the entire organization has set as a goal the conduct of more structured collaborative investigations into pregnancy outcomes occurring throughout both countries. There are currently two ongoing studies: the Asthma Medications in Pregnancy Study and the Rheumatoid Arthritis in Pregnancy Study. I'll describe the second in more detail.

The Rheumatoid Arthritis in Pregnancy Study is designed to evaluate the drug leflunomide. This drug was chosen because of concerns about the animal data and because its mechanism of action could be of concern in human pregnancy. OTIS member services throughout North America, in the course of receiving calls from pregnant women, refer those who might qualify for the study to a coordinating center. Each woman is asked to give informed consent to participate in one of three study groups: (1) those exposed to leflunomide, (2) those with rheumatoid arthritis not exposed to leflunomide, and (3) those who do not have rheumatoid arthritis and were not exposed to leflunomide. All women are interviewed periodically throughout and after completion of pregnancy, and all provide medical records from their health care providers and hospitals. Women with rheumatoid arthritis participate in a severity assessment of their disease, and women exposed to leflunomide have blood levels of the medication drawn early in pregnancy. Each child is examined before he or she reaches 6 months of age by one of four study dysmorphologists who are blinded to the study groups, providing a standardized evaluation for major and minor malformations in each child.

Strengths of this study design include (1) the precise collection of data throughout pregnancy on exposure timing, dose, and potential confounders; (2) engagement of the mother in the research effort, which minimizes loss to followup; and (3) most importantly, the intensive outcome evaluation that holds the potential for detecting a continuum of outcomes, including patterns of major and minor malformations, growth deficiency, and, to some extent, pregnancy loss. Finally, there is the potential to extend the study to evaluate long-term development of the children if that is a concern. Limitations of this study design include difficulty generating an adequate sample size to evaluate the risk of single major malformations with drugs not commonly used and the potential for bias from enrollment of a self-selected sample of pregnant women.

The final source of post-marketing information is birth defect surveillance activities. These include ongoing case-control studies such as the Slone Epidemiology Unit's Birth Defects and Environmental Exposure Study, which interviews mothers of malformed and nonmalformed infants regarding a variety of exposures during pregnancy. Similar in design are the state-based birth defects surveillance programs, some of which incorporate case-control studies that involve maternal exposure interviews. An example is the California Birth Defects Monitoring Program (CBDMP). Strengths of this monitoring program include (1) the use of active case ascertainment from multiple sources to identify malformed infants and stillborns, and terminated pregnancies with identified malformations; (2) the systematic selection of the control sample to be interviewed from a defined population; and (3) the ability to address associations with single major

malformations. The volume of information collected in maternal interviews allows the testing of multiple hypotheses and generation of new hypotheses over time, and the relatively easy incorporation of measures of genetic susceptibility and comparison of information across studies of similar design. Limitations of the CBDMP design include the retrospective collection of information on exposures and confounders and the voluntary nature of subject participation, both of which can introduce bias. And, finally, this approach does not allow examination of a continuum of effects beyond major malformations, such as spontaneous abortion, stillbirth, growth deficiency, and functional deficit.

So, considering all these various sources of information, why do we need to improve post-marketing surveillance? Consider six known human teratogens — warfarin, carbamazepine, captopril, valproate, misoprostol, and isotretinoin. The length of time between the marketing of these drugs and their first general recognition as teratogens ranged from approximately 3 years to more than 20. I think we would all agree this is less than optimal. Conversely, consider Bendectin®, which was marketed in the 1950s, used by 30 million women for the treatment of nausea and vomiting during pregnancy, and then withdrawn because of unfounded concerns about teratogenicity. The first major studies that actually assessed its teratogenicity did not come out until 17, 18, and 19 years after the drug was marketed, and multiple subsequent studies demonstrated overwhelmingly that the drug's margin of safety was probably within the acceptable range. If a systematic post-marketing surveillance program had existed at the time Bendectin® was first marketed, we might have avoided the drug's withdrawal, unnecessary anxiety for a huge number of families, and the litigation involved. Also, pregnant women might now have access to a relatively safe and effective medication to treat a common condition of pregnancy.

So, in summary, the use of medications during pregnancy is a very common event, it is very likely to occur early in pregnancy during what could be the most critical period of exposure, and there are limitations to all the methods in place today to determine the teratogenicity or relative safety of drugs. Our challenge is to try to improve these methods so we can shorten the time it takes to recognize that a drug might or might not be a problem.

References

1. Lacroix I, Damase-Michel C, Lapeyre-Mestre M, Montastruc JL. Prescription of drugs during pregnancy in France[letter]. Lancet 2000; 356(9243):1735–6.
2. O'Quinn S, Ephross SA, Williams V, Davis RL, Gutterman DL, Fox AW. Pregnancy and perinatal outcomes in migraineurs using sumatriptan: a prospective study. Arch Gynecol Obst 1999;263(1-2):7–12.
3. Forrest JD. Epidemiology of unintended pregnancy and contraceptive use. Am J Obst Gynecol 1994;170(5 Pt 2):1485–9.
4. Floyd RL, Decoufle P, Hungerford DW. Alcohol use prior to pregnancy recognition. Am J Prev Med 1999;17(2):101–7.

What Information We Need To Improve Post-Marketing Surveillance

J.M. Friedman, MD, PhD
Department of Medical Genetics
University of British Columbia

I think the information we need to improve postmarketing surveillance includes: (1) specifying the outcomes we must know about; (2) specifying the strength of the association, or the kind of risk, we are concerned about; and (3) identifying the practical issues and figuring out how to get the information we need. There is a long list of reproductive outcomes that are of interest. Most of the surveillance systems we now use deal with structural congenital anomalies, particularly major malformations, but we also need to be concerned about syndromes and more subtle patterns of congenital anomalies. Distinctive patterns of congenital anomalies, such as fetal alcohol syndrome, are often more sensitive than major malformations as indicators of teratogenic effects. There is also some question about whether we should be looking at minor anomalies, which are more difficult to define consistently and more problematic to interpret. Functional abnormalities can also have their onset before birth. Mental retardation, deafness, and

blindness are examples. Learning disabilities are a frequent outcome and therefore of concern to many people, and autism is a very serious disorder that is increasingly in the news. Functional disorders that do not involve the central nervous system can also occur, such as the renal impairment associated with captopril use during pregnancy.

A variety of other outcomes are sometimes studied, including premature delivery, spontaneous abortion, late fetal death, and stillbirth. Death in infancy or later childhood would also be of concern if we knew there were drugs that caused this. Chromosomal abnormalities or new mutations are important outcomes, and there are a whole variety of new mutations, including autosomal dominant, autosomal recessive, X-linked recessive, and mitochondrial mutations, that could be evaluated. Measures such as birth weight, birth length, head circumference, and growth during childhood can be concerns. The experience with diethylstilbestrol (DES) taught us that transplacental carcinogenesis does occur in humans, and this is also an important concern. There are claims that other adult-onset diseases, such as heart disease or psychiatric illnesses, could be associated with prenatal exposures. Second generation reproductive effects would also be important to know about if they occurred.

The list is actually much longer, as there could be effects on fertility and reproductive function for the mother before she conceives, or for the father. We are interested in all of these outcomes, but there is no practical way to study them all. We have to find a practical way of conducting surveillance that will yield a signal when there is a medication that really is causing a problem about which we should be concerned. So, the question really should be, "What risks related to maternal medication treatment during pregnancy are essential for us to know about?"

Factors that can help decide this include the severity and frequency of the outcome, and the associated practical issues. For example, a relationship between a maternal treatment during pregnancy and coronary artery disease in the offspring might be very important because coronary artery disease occurs so frequently. But we are not likely to identify this association through a surveillance program because the outcome does not occur until 50 years after the pregnancy exposure.

In terms of severity, conditions that are life threatening or lethal, that are severely handicapping, or that are irreversible or untreatable would be priorities. Examples include (1) phocomelia from thalidomide, which results in serious physical handicaps; (2) fetal alcohol syndrome, in which there are serious behavioral and neurocognitive handicaps; (3) isotretinoin embryopathy, which results in serious physical and neurocognitive defects that are often lethal; and (4) captopril, with which the main adverse outcome is fetal death from oliguria and renal failure. At the other end of the spectrum, conditions that could be considered of mild severity include staining of the primary dentition from tetracycline and transient neonatal hypertrichosis with minoxidil. These have no functional significance and disappear with time, so would be fairly low on the list of concerns.

The three chief ways to quantify frequency are with absolute risk, relative risk, and population attributable risk (population attributable fraction). Absolute risk is defined as the risk that a woman who takes a particular medication during pregnancy will have an affected baby. "If you take this medication during pregnancy, there is a 10% chance your baby will have that outcome," for example. Relative risk expresses how much more likely a woman who takes a particular drug during pregnancy is to have an affected baby compared with a woman who does not take the drug. If you have a 15% risk of miscarriage with this medication, but that is the same risk as someone who does not take the medication, there is probably no need to be concerned. Population attributable risk is the proportion of adverse outcomes of a particular type in the population as a whole that are caused by use of a particular drug during pregnancy.

There is an interesting relationship between population attributable risk and relative risk. Exposure to a particular drug often occurs in less than 1% of all pregnancies in the population. When the exposure rate in a population is this infrequent, an enormous relative risk (on the order of 10 to 100) is needed to have an appreciable effect on the fraction of infants in the population born with the condition. Similarly, for many exposures that do

have an adverse reproductive effect the associated relative risk is low, in the range of 1.5 to 3.0. When the relative risk is this low, a very high exposure rate (on the order of 10% or more) is needed to affect the population attributable fraction substantially. This is one of the practical problems that must be dealt with.

This leads to the issue of study power, which really shapes everything we do. Power can be defined as the chance of finding an association that really exists. It is a statistical way of saying how likely we are to pick up what we are looking for if it is really there. Power depends on several aspects of a study, including the sample size, the frequency of treatment with the drug, the frequency of the outcome in the population, and the strength of the association between treatment and outcome. For example, we might be interested in evaluating whether a drug causes a twofold increase in mental retardation. First, we assume that the frequency of treatment with this drug is 1% and that the expected frequency of severe mental retardation in the general population is about 1%. We specify a, the usual statistical significance cutoff for the P value, at 0.05, and specify 1-b, which is our power, at 0.80. This means there is an 80% chance that, if there really is an effect, we will pick it up at that level of statistical significance. If we conducted a cohort study, we would need to observe 1,116,000 children to identify a twofold increase with 80% probability. If we conducted an exposure cohort or case-control study and there was a one-to-one match between the exposed and the unexposed groups, we would need just over 5,000 children to identify this association. The issue of power is a real problem in reproductive toxicology studies because they very often deal with rare treatments, particularly if limited to a particular point or critical time in pregnancy. They also often deal with rare outcomes. The outcome usually is not all birth defects, but one particular defect, one particular functional outcome, or one type of cancer that occurs later in life. To make matters even more difficult, reproductive toxicology studies often deal with weak associations, with relative risks in the range of 2 or 3. In general, enormous numbers of study subjects are needed to detect increases in this range. To detect something with a relative risk of 20, the numbers do not have to be so big, but it is very uncommon to see a relative risk of 25 or 10 or even 5 in these studies. We have to come to grips with how big a difference from what is expected in the general population we are willing to tolerate if our surveillance is to be reasonable.

There are additional practical issues to be considered when thinking about methods for post-marketing surveillance for the effects of medications in pregnancy. One is the difficulty of diagnosing the outcome of interest. Some outcomes are very easy to diagnose while others are much more difficult. Neonatal death is pretty easy, as is anencephaly, although the latter isn't always recorded accurately on birth certificates. Congenital heart defects are more difficult to diagnose and often require an echocardiogram or evaluation by an expert. The diagnosis of fetal alcohol syndrome usually requires evaluation by a pediatrician familiar with this pattern of anomalies, and there is disagreement on how to define autism. The age at diagnosis can also vary greatly from something like anencephaly that is very obvious in the newborn, to something like congenital heart disease that is often picked up later in infancy, to learning disabilities that might not become apparent until a child is 3 or 5 years old. Some outcomes might not be evident until adulthood, for example, the cancer seen with prenatal DES exposure.

Both the difficulty of diagnosis and the age at diagnosis relate to the cost of obtaining data. We live in a world of real resources and we probably can't afford to know everything we want to know about the effects of drugs on the developing fetus. One of the most important factors in the cost of obtaining data is whether the data already exist. If data have already been collected that can be used if put in order, that usually is a more efficient way to do things. For example, information on neonatal death is collected by all jurisdictions, states are beginning to collect data on congenital deafness, and a number of states have registries that collect information on major malformations in a standard format. But there is the additional issue of linking these existing data to the mother's exposure during pregnancy, and this is not a trivial issue. More expensive are data collected specifically to look at pregnancy outcomes. Getting information from existing hospital and physician records is usually much less expensive than obtaining it from maternal

interviews, but certain questions cannot be answered from routine records. Examination by a specialist, such as a dysmorphologist, might not be expensive to do but arranging such examinations as part of a surveillance system would not be easy. If special tests such as echocardiograms or magnetic resonance imaging (MRI) are required to make a particular diagnosis, the data are much more costly and much less reasonable to obtain as part of a surveillance system.

So, what information do we need to improve post-marketing surveillance? We need to specify the adverse outcomes and the magnitude of risks that are most important to know about, and we must be practical in making our choices. I would suggest, for example, that we want to be sure that no drug produces a teratogenic effect equivalent to that of thalidomide or alcohol embryopathy without our knowing about it. In order to do this, we would have to be able to detect at least a twofold increase in major malformations, a tenfold increase in mental retardation, or a unique syndrome in at least 10% of the offspring. Once we have made these choices, we have to figure out a practical way to gain this information.

How To Improve Post-Marketing Surveillance: Constraints We Have To Live With — Methodological Issues

Margaret Honein, PhD, MPH
Birth Defects and Genetic Diseases Branch
National Center for Environmental Health
Centers for Disease Control and Prevention (CDC)

I will present some of the methodologic challenges and constraints inherent in trying to study the safety or risk of prescription medication use in pregnancy and how they affect different study approaches to greater or lesser degrees and in varying ways. An increasingly challenging problem for all studies is the issue of participation. The ability to make unbiased inferences about exposures ideally requires recruitment of an unbiased sample of pregnant women who have had the exposure of interest and a comparison group that is identical except for that exposure. At the least, the comparison group should differ from the exposed only on confounders that can be measured and controlled for in the analysis. However, what often happens is that a study either (1) recruits a small sample of women with exposed pregnancies who are referred by their providers or self-reported without an adequate comparison group, or (2) recruits an appropriate comparison group but enrolls too few women with exposed pregnancies to detect small to moderate risks.

Recruiting study participants from health care providers might be an efficient means of enrolling women with high-risk pregnancies and specific drug exposures, especially if one or a relatively small number of large clinics manage these patients. However, the challenge of recruiting a sufficiently large sample of women remains. Health care providers might not perceive a benefit, or actually might perceive some risk related to patient confidentiality, in reporting exposures. Given a mother's underlying medical condition, the health care provider might view the exposure as necessary and not something particularly unusual or risky. Reporting patients to registries and studies also requires time, and health care providers increasingly lack sufficient time to take on additional tasks.

Another strategy for enrolling participants is to directly recruit pregnant women exposed to medications. However, it is very difficult to identify and enroll these women early in pregnancy. Those that do enroll probably do not represent a random sample of all exposed pregnant women, as enrollment usually is an active process that requires completing forms, returning phone messages, providing informed consent, and such. Many passive refusals can occur through lack of response. In addition, society today is very mobile. A recent case-control study estimated that approximately 20% of women moved between the beginning of pregnancy and the date of delivery. It is not possible to follow all of these women throughout their pregnancies.

Recruiting patients into traditional case-control or cohort studies has also become more challenging. There are new telephone features, such as caller ID, that make it easier for people to screen calls and avoid answering them. Also, the public today is

probably more sensitized to issues of privacy and whether answers are kept confidential, and perhaps is less trustful of the scientific community and less eager to participate in studies, than in previous years. There can be particular sensitivity about genetic testing and similar issues. And, in general, people perhaps have more constraints on their time now.

Promotional efforts can increase participation in studies but, depending on the type of study, there are constraints on which efforts can be used. For example, industry representatives might not be allowed to promote registry enrollment directly to physicians because the drugs being monitored might not be licensed for use during pregnancy. Promoting the registry might also be viewed as promoting use of the drug during pregnancy. Displays and presentations at professional meetings and articles in journals can make health care providers aware of the existence of registries and the practical methods for reporting exposures during pregnancy. However, to be effective, promotional efforts usually need to be an extensive and continual effort.

Exposure cohort studies that enroll women directly require promotional efforts to a wider group of women. It is important to keep health care providers in the loop, as they can encourage the patients they see to enroll. Placing brochures about studies in their offices is one method to promote enrollment. However, direct promotional efforts, such as placing articles and advertisements in the popular press, are needed to reach a wider group of women. And, even with all of these efforts, reliance on voluntary reporting can still result in substantial selection bias.

Regardless of the type of study undertaken, a substantial infrastructure will be needed. All studies require personnel to perform the data collection and database management, funding sources, and a mechanism to disseminate findings to the people who need them. For example, a single drug registry would employ an epidemiologist, project manager, database manager, programmer, statistician, office equipment, and likely some sort of advisory committee to oversee the project. The estimated cost is maybe $200,000 to $300,000 per year. A birth defects case-control study would require even greater infrastructure. Many such studies ascertain cases from an existing birth defects surveillance system, which has its own extensive infrastructure. In addition, the study would employ clinicians to review the cases, epidemiologists, a project manager, database managers, interviewers, computers, office equipment, and such. The National Birth Defects Prevention Study, a case-control study of children with major birth defects, currently interviews the mothers of over 3,000 case and control children combined each year. This requires extensive resources in addition to the ongoing costs of the surveillance program.

Currently in the planning stages is a longitudinal cohort study of children's health in the United States which would track women and their children over long periods of time. This will require substantial personnel and resources but has the potential for tremendous benefit in what can be learned. However, even an expensive, extensive effort like this has constraints in addressing the specific issue of post-marketing surveillance for prescription drug use during pregnancy. The exposures and outcomes of interest are likely to be quite rare, even in a very large cohort, and because new drugs continually enter the market, what is learned at one point in time might not be sufficient for what is needed 5 or 10 years later.

There are also a number of important issues related to data quality: (1) the validity of the exposure information, (2) the validity and completeness of the outcome information, (3) the time during gestation of enrollment in the study, (4) the availability of information on confounding variables, and (5) the availability of a comparison group. Concerning the validity of exposure information, the ideal is to know the dose, duration, and timing of the drug exposure during pregnancy. This information is usually obtained through maternal reports, but can be verified with pharmacy or medical records in some study designs. Most inadvertent exposures occur in the first trimester, before the woman realizes she is pregnant. However, drugs used to treat underlying maternal conditions might be needed throughout pregnancy, so that it is difficult to find a group of women who were exposed at a single point in their pregnancy.

Single drug registries ascertain exposures reasonably well as they occur, and a longitudinal

cohort study presumably would do the same through maternal reports. Case-control studies obtain information about exposures months to years after they occur, resulting in at least the potential for recall bias. Some exposures might be substantially underascertained after the completion of pregnancy compared with those ascertained during pregnancy. Examples are the use of over-the-counter medications, episodic occurrences such as fever, and sporadic use of prescription medications.

Complete outcome information can be very difficult to obtain. Reports with information noted only by the obstetrician immediately after birth, as are obtained by many single drug registries, can be incomplete, particularly if a full pediatric examination is not included. Obtaining medical record documentation of the defects is helpful, but an examination by a dysmorphologist could be required to ascertain patterns of minor malformations such as those seen with fetal alcohol syndrome. In addition, prenatal diagnosis and termination of affected pregnancies can differ among exposed and unexposed pregnancies. If the outcomes included in a study are only from among live births, the results might provide a very incomplete picture of the outcomes that occur following exposure.

There are two important issues related to the timing of enrollment into exposure cohorts. The first has to do with spontaneous abortion, which is a relatively frequent adverse reproductive event. A higher rate of spontaneous abortion in women exposed to a drug might suggest teratogenicity. However, because the frequency of spontaneous abortion declines throughout pregnancy, the ability to evaluate this outcome is dependent on enrolling women relatively early during their pregnancy. It is not necessary to enroll all women at the same gestational age, this can be controlled for in the analysis. But, for example, if a substantial proportion of the women are enrolled after 20 weeks gestation when the frequency of spontaneous abortion is relatively low, it will not be possible to accurately assess the frequency of this outcome.

The other issue related to the timing of enrollment is the exclusion of women from exposure cohorts who have already had a fetal defect diagnosed prior to enrollment in the study. For example, a woman who has had an ultrasound at 12 or 15 weeks that showed a neural tube defect would not be included in a prospective cohort. However, a woman who has had a normal ultrasound, even late in pregnancy, could be included. This cold lead to an artificially low baseline prevalence of birth defects in the study. For example, if an exposure cohort of women enrolled after having a normal ultrasound is followed and 3% of the women have an infant with a birth defect, this might actually represent an increase in the frequency of defects.

To adequately assess the relationship between an exposure and an outcome, accurate information on confounding variables such as smoking, diet, genetics, and illicit drug use is needed. Most cohort and case-control studies ask women this information and, to the extent that it is accurately reported, are able to control for these factors in the analysis. However, the accuracy of this type of information varies widely. Smoking behavior is fairly well-reported, but alcohol use is poorly reported in most studies that have assessed validity. Additionally, some of the current exposure cohort studies ascertain information on confounding variables only among women who have had an adverse outcome, not those with normal outcomes. This makes it difficult to evaluate the impact of confounding on the relationship between exposure and outcome.

The availability of an appropriate control group is another important methodological issue. Ideally, an unexposed group as comparable as possible to the exposed group is needed to compare experiences. When it is not possible to enroll a control group, alternatives include comparing women exposed to the drug of interest with women exposed to other drugs or to drugs known not to be teratogenic. Another option is to compare women exposed to the drug of interest during the first trimester with those exposed only during the second or third trimesters.

Given these many data quality issues, interpreting the findings of each type of post-market study is difficult. This makes it even more important that the appearance of a conflict of interest in these interpretations be avoided. Both the credibility of reported findings and the willingness of health care providers to report exposed pregnancies might be improved by ensuring that scientific independence exists. Some industry-based registries have

addressed this issue by assembling an independent scientific advisory committee to interpret the findings.

In conclusion, better information on the safety of prescription medication use in pregnancy is needed. While the existing constraints and challenges are important to consider as we discuss these issues, they should not be seen as insurmountable problems.

How To Improve Post-Marketing Surveillance: Constraints We Have To Live With — Ethical and Legal Issues

Zita Lazzarini, JD, MPH
Director, Program in Medical Humanities, Health Law, and Ethics
University of Connecticut Health Center

The ethical and legal issues involved in long-term monitoring of the use of medications in pregnancy can be broken down into three basic areas: (1) the public health justification for monitoring, (2) the ethical issues related to monitoring that include both justifications and obligations, and (3) the laws that get tangled up in long-term monitoring. From these, it is possible to put together some potential criteria necessary for a surveillance or monitoring program to meet minimum ethical and legal standards. The issues related to monitoring the use of antiretroviral drugs in pregnancy provide a helpful example. While relatively few women take antiretroviral drugs during pregnancy, there is very strong ethical justification for monitoring these exposures based on the way the drugs are prescribed to women, and there are other ways that the issues in monitoring antiretrovirals are both similar and dissimilar to those involved in monitoring other classes of drugs. Antiretrovirals are both easier and more difficult to monitor, and understanding the experience with them could be useful in other settings.

The public health justification for long-term monitoring of potential toxicities resulting from the use of antiretrovirals during pregnancy lies in the fact that their use has made prevention or reduction of the rate of transmission of virus from mother to child both a possible and a desirable goal. This has led clinicians and public health officials to strongly encourage women to be tested and to accept treatment if they are infected. This requires talking with women and counseling them, and this is not entirely nondirective counseling. The process of obtaining informed consent actually verges on suggesting that women take these drugs. This might be similar to other situations in which there are distinct benefits of continuing treatment for a chronic illness during pregnancy. In the case of antiretroviral drugs, there is clearly a public health benefit in reducing HIV transmission to children and, in some cases, there is benefit to the women to be treated during pregnancy. However, this treatment might also expose the women to some long-term risks and, more particularly, might expose their fetuses to a small but unknown risk. Long-term followup is needed either to define this risk or to rule it out to be able to fully inform women in the future.

Ethical issues related to monitoring include both justification and obligation. The consequentialist approach suggests that if the benefits of long-term monitoring outweigh the risks there might be an actual obligation on the part of clinicians, pharmaceutical companies, and public health officials to aggressively pursue long-term monitoring. In this case, the benefits would include early detection of adverse events, dissemination of that information, and protection of future groups of women and babies. The risks mostly focus on the confidentiality of health information, although there could be others. Monitoring systems already in place in this area include the original clinical trials, post-marketing surveillance, public health initiatives to use existing databases where they exist, HIV registries, birth defects and tumor registries, and special purpose registries such as the antiretroviral pregnancy registry.

Another ethical perspective is the principalist approach. This suggests there are fundamental ethical principles that require health authorities to avoid or mitigate risks related to recommended treatments and that in many ways support long-term monitoring of the antiretrovirals and probably many other classes of drugs that women take. The principle of beneficence states there is an obligation to choose the best treatment and to do the best by your

patients. This suggests that physicians who have patients with chronic diseases will want them to have those conditions treated. In the case of HIV-infected mothers, it suggests that taking antiretrovirals while pregnant will generally benefit their children because it so reduces the risk of transmission. The principle of autonomy is extremely important, mostly in the sense that data from long-term monitoring are really needed for women to make informed choices about whether to take antiretroviral drugs during pregnancy. This is even more applicable in other situations where alternative therapies are available and the decision revolves around choosing which therapy to take.

The principle of nonmaleficence means to do no harm. This applies to those who suggest to pregnant women that they take medications during pregnancy, those who make these drugs available, and those who market the drugs, all of which make it likely that large numbers of women of reproductive age will end up taking them. It suggests that both short-term adverse events and long-term harms should be detected as soon as possible so that recommendations and counseling about the use of these drugs can be changed when necessary. Lastly, the principle of justice suggests that one of the ways to potentially mitigate burdens is to detect risks early and prevent future cohorts from being exposed, if possible.

So, based on the public health justification and these ethical approaches, we can form a list of basic obligations for the area of conducting long-term monitoring. A fundamental place to start is with the obligation to obtain informed consent for treatment. There is also an obligation to inform patients about the collection of data. If long-term monitoring is going to take place, patients should know that information about their birth outcomes will be monitored and whether there are other proposals in the system to interview them or collect data on them. There is an obligation to provide followup for adverse events and outcomes of pregnancy throughout the duration of the individual relationship. This might be the relationship between physician and patient, which is probably the shortest, or it can be thought of more broadly in terms of the pharmaceutical company's relationship with clients who have taken its drugs, and, beyond that, to the relationship between public health and the public in general. There is also an obligation to prevent or mitigate harm that might have been caused by medications that are either made available or recommended, and to collect, store, and use data on adverse outcomes to maximally benefit both individuals and the public health.

The distinction between ethical and legal issues is always somewhat tricky, as a number of issues can fall into either category. In the ethical area are basic principles and approaches to problems and the normative goals of what should be done by clinicians, health care providers, and public health officials with some reference to professional codes of ethics. The legal area is the part that some people call "ethics with teeth." It is where the law has stepped in to grant authority to do something, to set limits on what can be done, or to act as a tool to regulate or change behavior. The latter can apply to individual providers, to the public health system, or to individual patients. Issues that are covered by both ethics and law include those of consent, notification, confidentiality, and the protection of patients from risks and harm. They appear in both because a large part of our regulatory systems have been based on what people thought were ethical duties or obligations.

Now, what role does the law play in addition to ethics? It grants authority, sets limits, provides protections, and establishes penalties and immunities. But in some ways, the law can make things more confusing. For example, under the law, similar information about a single patient can be treated differently depending on who holds the information. There can be a clear ethical duty to conduct followup after an exposure, and yet there can be conflicting legal standards about what you are allowed to do. The law can create barriers to sharing information, and it is always impressive to see what sharing is actually done beyond what it appears might be possible. The law can also set minimum standards for the collection, use, and protection of information. This is the most important part in relation to developing future activities.

So, how do legal issues relate to monitoring the long-term effects of antiretrovirals and other classes of drugs? There are a number of legal principles, or areas of law, that can and will intrude upon this process. These include fair information practices;

the distinction between public and private; the distinction between research and surveillance; and the state, federal, and model laws and regulations that govern health information. Additional interesting issues are those related to sharing information across databases and potential liability. Fair information practices came out of a report in the early 1970s from the U.S. Department of Health, Education, and Welfare that set up principles for the collection and use of data about individuals that are particularly relevant in the health area. These have been continued in more or less the same form through most of the recommendations that have come through the system in the last couple of decades, including the new Health Insurance Portability and Accountability Act (HIPAA). The basic principle is that people have a right to know when information is going to be collected about them. They have a right to be notified in writing, to access the information, to inspect it, to copy it, to amend it if necessary, or to add corrections to it. They also have the right to an accounting of the disclosures that happen, sort of an audit trail. And, there is a corresponding responsibility on the part of those who collect the information to protect it and to establish technical and procedural security measures to make all of these things possible.

The public versus private distinction becomes particularly important when the public health system is collaborating with the private sector. The boundaries between who is acting in the public sector and who is acting in the private sector can sometimes get fuzzy. What we might need to do is make any combined system sufficiently protective to meet the higher standards of both. But in the health area, public health information is pretty clearly defined as that held and collected by government agencies, and private health information is that held by everyone else. In fact, the same information about an individual can be included in both areas, depending on who holds it, which is tremendously confusing in practice. This public versus private distinction has a definite impact on the requirements for oversight and for whether consent is needed, that is, whether participation in the data collection is voluntary. It also affects how the data are used and reused, and potentially whether the information is redisclosed.

The distinction between research and surveillance is also important. Government surveillance is usually based on some type of mandate to collect the information, which implies a whole different range of powers and obligations compared with other types of data collection and research. Because research is thought of as potentially putting people at risk, regardless of whether they also benefit, a range of regulations and protections have been imposed on the use of human subjects in research. This makes it more complicated. The need to go to women to obtain their consent to participate in long-term monitoring that is part of research can actually complicate the process of collecting data. But, it is thought to be necessary on a societal level to protect people from the small but real risk undertaken when they participate in research. The Code of Federal Regulations defines research as a systematic investigation designed with the intent to develop or contribute to generalizable knowledge. It defines a human subject as a living individual about whom a researcher obtains data either through intervention or interaction with the individual, or from identifiable private information. The distinction between surveillance and research affects what is necessary for oversight, whether the participation is voluntary, whether consent is needed, and the types of uses and reuses of the data that are allowed. If an activity is research, ethical review will be required to weigh the benefits and burdens to each subject. The subjects will need to be informed and their voluntary participation ensured in almost all situations. Mandated surveillance systems run by government agencies do not require any of those pieces.

Which laws and regulations govern the collection of data by a system set up to follow women during pregnancy will depend on several things. If the system conducts research involving human subjects, then the whole range of human subjects regulations will apply. At the federal level, the federal privacy laws and regulations will apply, to different degrees. At the state level, there is a range of potential provisions including public health laws governing the conduct of surveillance; public health data laws governing how data are protected and shared; health care information laws that are much less extensive than data laws; and tort law that governs the potential for liability, both medical malpractice

liability and product liability. A survey conducted during 1995 and 1996 found that public health data laws among states across the country were consistent in that all states protect public health data. But there was substantial variation from state to state regarding which data are protected and how. This is important because, when using data collected by the government, the particular ins and outs of how those data can be used or shared and the need for rejustification for secondary uses of the data can vary among states. This can create barriers to the interstate transfer of data and to its use by private entities for purposes not closely related to the original purpose for which the data were collected. In addition, while health care information laws mostly cover the doctor-patient relationship, a few states have extended these to cover insurers, researchers, educators, and other employers and entities that hold health care information.

It is important to be familiar with the State Model Public Health Privacy Law. This is a model law, meaning that it has no force unless a state adopts it. But if adopted by states, it would set more uniform criteria for the collection, protection, and use of publicly held data, which could be very beneficial in attempting to conduct widespread, or perhaps nationwide, monitoring of pregnancies. The Model Law does not pose absolute barriers to this type of monitoring, as it allows the use of data for public health and research purposes. What it does is define various terms, including what is meant by a public health agency, protected health information, legitimate public health purposes, and fair information practices. It also establishes conditions for the acquisition, use, and storage of information; sets criteria for the disclosure of information with and without consent of the individual; and limits secondary uses of the data, among other things. Basically, it allows the acquisition of information that is directly related to a legitimate public health purpose, is reasonably likely to achieve that purpose, and cannot otherwise be achieved with non-identifiable information. Use of such data is limited to legitimate public health purposes directly related to the purpose for which the information was acquired. This may allow the use of person-identifiable information if that information is necessary to monitor and investigate drug toxicities.

There are also existing federal laws and regulations that apply, although federal information privacy provisions really apply only to federal agencies. There are federal regulations concerning special subject areas, such as substance abuse, that are not directly related to medication use during pregnancy. Also, there are research certificates of confidentiality that are important and can provide added protection when setting up a research protocol. But, probably the most significant are the federal rules put out as a consequence of HIPAA. These do not apply to all holders of health information, but they do reach a substantial number of those holders and set boundaries on the collection and definition of types of health information. They include security requirements for those who handle health information and provide for consumer control. The latter goes back to the fair information practices idea that people have the right to be informed of, to inspect, to copy, and to amend or correct information that is collected about them. The rules provide for accountability and public responsibility on the part of those who hold health information. They also state explicitly that the interests of individuals and their privacy shall not completely outweigh those of the public in certain uses of health information, and that is a very important piece that sometimes gets lost. They state that, not only does privacy have to be protected, but the important public uses of health information collected about people must be considered. Of the many uses of health information that are permitted without the individual's consent under these rules, those relevant to long-term monitoring include public health functions, research, perhaps emergency circumstances, the provision of information to the next of kin and to a government health data system, and as otherwise required by law. So, while these rules set some frameworks and some necessary constraints, they do not appear prohibitive to setting up either small or large systems of long-term monitoring. They have substantial gaps in coverage, in that they do not reach many kinds of industry and researchers. But their intent is to provide a basis for the public uses of information and for protection of that information.

The primary question, really, is: Do current laws require consent, confidentiality, or notification of women to participate in any kind of monitoring system? And, unfortunately, the answer is that it often

depends on some finer distinctions or perhaps future interpretations. But it breaks down into some of those things already mentioned: Is the activity research, or is it surveillance? Is it a public activity, or a private one? Clearly, if it is research and it is federally funded, then notification, consent, and confidentiality protections are required. That is the most protected and the most regulated situation. If the activity is research based solely on private funding, whether notification, consent, and confidentiality are required probably depends on how the data are to be used. If the activity is used to submit data to the Food and Drug Administration, then requirements very similar to those for federally funded research apply. If data are not submitted to a federal agency, the requirements might be much less, as is now the situation for some pharmaceutical industry activities.

If the activity is surveillance conducted by the government, notification is still supposedly required, though in reality many people might not actually know when surveillance about their reportable diseases and other conditions takes place. And that says that we often do a poor job of letting people know some of the public health functions that are ongoing. However, notification is not the same thing as consent. Many kinds of surveillance mandated by the government do not require consent. Nonetheless, consent might be required for a new surveillance system that involves more intrusive activities, such as interviews or in-depth investigation. Confidentiality is required for surveillance activities, but there are areas for which specific disclosures are permitted, such as in communicable disease surveillance. Whether confidentiality is required for surveillance conducted by a private entity really depends on what the data are to be used for, who the private entity is working with, who the subjects are, and other things. In such cases, there is probably an ethical standard that is a little higher than the legal standard.

Given all the laws that are in existence, how do we go about using health information? It might be that we want to share data from different databases. Potential issues include the compatibility and format of the data collected, and the confidentiality and data use laws, regulations, and policies of the particular areas where the databases already exist. As an illustration, consider three states (California, Connecticut, and New York) and three types of data bases (tumor registries, birth defect surveillance programs, and the Antiretroviral Pregnancy Registry that is sponsored by multiple pharmaceutical companies and includes both public and private providers). The tumor registries in all three states make all information that is collected confidential. The Antiretroviral Pregnancy Registry, on the other hand, uses linked but anonymous data where the health care provider keeps the link to the woman's identification. The Connecticut Birth Defects Registry provision does not mention confidentiality. It is part of the public health system and is probably protected under broader public health regulations. The policies of other states are unknown, and there are probably a variety of approaches. As far as data sharing goes, the tumor registries in California, Connecticut, and New York mandate reporting, including personally identifiable information. They permit inspection of the information and allow data sharing out of state on the condition that certain types of confidentiality protections are ensured. However, the requirements for ensuring those protections vary among the three states. Some require a designation by the commissioner of health or the head of the registry, others require that the collaborating state have certain provisions in place. Only New York specifically addresses the question of standardizing the information so that the data system is compatible with others. So, if the tumor registries in California and Connecticut wanted to compare their personally identifiable data with the anonymous data from the Antiretroviral Pregnancy Registry, they would need to find a way to link those without destroying the confidentiality protections already in place for the registry data. A similar problem would likely occur in any state that has an HIV reporting system that does not include names, for example, and wants to link that data with birth defects or tumor registries.

The last legal topic to be considered is the potential for liability. In reality, liability is a minuscule part of the area of public health and should remain so. Liability is certainly a consideration, but it should not displace public health benefit, individual benefit, good clinical outcomes, and other factors as the driving force behind long-term monitoring. In the circumstance in which risks are known before a drug is approved or before treatment is

recommended, then liability is likely if it is later determined through post-marketing surveillance that complications have occurred and that women were not told about the potential risk. But, in the circumstance in which the risks are not known or in which women were informed of the level of known risk, there are a number of ways that the potential for liability can be decreased. Informed consent and fair disclosure of known and potential risks are important. Continual monitoring for any potential risk is also important, as is a prompt response to any identified problems. These are part of the public health justification for monitoring. It is possible that long-term monitoring could increase a company's risk of liability if previously undetected problems were uncovered. An extensive monitoring system would pick up risks not otherwise recognized. But, the much greater potential for liability actually comes in the failure to obtain informed consent, failure to monitor, and failure to report premarketing adverse events or to diligently follow up on ones that occur.

So, the minimum standard of what is needed to facilitate long-term monitoring that is both ethical and legal includes the following. The first element is to have consensus on the ethical obligations involved, that is, how and why we are collectively responsible for long-term monitoring. The next element is the voluntary participation of clinicians, pharmaceutical companies, public health officials, and probably women, all of those who participate in one way or another. This is important in terms of recruitment and in terms of educating people. Next is to require coordination of all systems set up so that they meet minimum ethical and legal standards, both state and federal laws, and to identify practical barriers to meeting these standards on an ongoing basis. Sometimes the barriers are legal, sometimes they are simply things that happen in practice, but they can interfere with long-term monitoring and could increase the likelihood that things will be conducted in an unethical or less than legal way. Identifying them early and trying to find ways to solve them is probably the best approach.

Related Literature

Gostin LO, Lazzarini Z, Neslund VS, Osterholm MT. The Public Health Information Infrastructure. JAMA 1996;275:1921–7.

Gostin LO, Lazzarini Z, Flaherty. Legislative Survey of State Confidentiality Laws, with Specific Emphasis on HIV and Immunization. Final Report Presented to the Centers for Disease Control and Prevention (1997).

Health Insurance Portability and Accountability Act (HIPAA) [online]. 1996. [cited 2000 Nov 29]. Available from URL: http://aspe.os.dhhs.gov/admnsimp.

Model State Public Health Privacy Act [online]. 1999. [cited 2000 Nov 29]. Available from URL: http://www.critpath.org/msphpa/privacy.htm.

Advisory Committee on Automated Personal Data Systems. Records, Computers, and the Rights of Citizens. Washington, DC: US Department of Health, Education and Welfare; 1973.

How To Improve Post-Marketing Surveillance: Constraints We Have To Live With — Use and Linkage of Large Data Sets

Wayne Ray, PhD
Professor and Director, Division of Pharmacoepidemiology
Department of Preventive Medicine
Vanderbilt University School of Medicine

The study of drug effects on the fetus is part of a considerable public health problem in this country in that a substantial number of people are exposed to medications, the effects of which are really unknown in many important aspects. Under the leadership of the Agency for Healthcare Research and Quality, the Centers for Education and Research on Therapeutics have been established as a cooperative effort to provide a better infrastructure for learning what we need to know about medications and communicating that knowledge to practitioners. Part of that infrastructure includes the use of automated databases in the study of fetal drug effects. An automated database can generally be defined as computerized records of medical care provided to a defined population. At an operational level, these computerized records are usually defined in terms of specific files that have a physical existence but also reflect important functionality. The enrollment file describes the defined population

that is in the system. The pharmacy file describes medications that the patients presumably receive. Sometimes these contain records of actual prescriptions filled at the pharmacy; sometimes they contain records of prescriptions written by physicians that were intended to be filled and taken; and sometimes they contain the actual medication administration records from health care settings, such as hospitals. Regardless of the specific type, the pharmacy file contains some kind of automated record of medications that patients are to receive and presumably take. A third file contains computerized records of medical encounters, such as hospitalizations, physician visits, emergency room visits, and stays in nursing homes.

Some of the earliest automated databases were those from health maintenance organizations (HMOs), such as Group Health Cooperative of Puget Sound and the Kaisers of Northern and Southern California. Here, the defined population was the members of the HMO about whom computerized records in each of these areas were kept for a variety of reasons. Medicaid programs that provided medical care to the qualifying poor also kept this kind of information for somewhat similar reasons. And in the United Kingdom, there is now the General Practice Research Database that has grown out of the computerized medical records systems.

The following examples are based on experience with the Tennessee Medicaid program. Briefly, Vanderbilt University has a partnership with the Tennessee Health Department, where the Medicaid program is based. Currently, the base Tennessee Medicaid population includes 1.4 million people and about 30,000 births per year, or about half of the births in Tennessee. As part of a national effort to improve perinatal outcomes, Medicaid has been used as a vehicle to ensure delivery of care throughout pregnancy and then to young children. And so, throughout the country, the birth rate in the Medicaid population is higher than in the general population. The database contains longitudinal data from 1974 and a fair amount of experience with it has been accumulated, particularly in the area of linkage with vital records, which is quite important for studying fetal effects.

Okay, why use automated databases? There are many logistical obstacles facing researchers in this area. But the huge advantages of databases are, first, that they provide the kind of defined population needed to study exposed and unexposed pregnancies. This is true regardless of whether this includes members of an HMO, people in a Medicaid program, or those in a computerized medical records system. A second advantage is that the history of medication use during pregnancy can be reconstructed from the pharmacy files, if conditions are right. This history is collected for the entire population and is collected prior to the birth outcomes, so that it is not biased by the outcome except in very indirect and unusual ways.

One example of how hard it can be to get this information using traditional methods can be seen from an interesting study done in Europe on the validity of interview data for measuring drug exposure during pregnancy. In this study, data were collected on 488 high-risk pregnancies at two time points: once during pregnancy from the physicians treating the women, and then again 7 years later. The idea was that the later data collection would, perhaps, emulate a large case-control study of birth defects that had occurred over many years. What was found was that only 55% of the drugs administered in pregnancy were reported in the interviews 7 years later, and only one-third of the drugs received in the first trimester were reported 7 years later. This shows the formidable difficulties faced in obtaining drug information from interviews. In addition, as known from some of the early studies of Bendectin® and congenital heart anomalies, not only is information about medication exposure during pregnancy likely to be incomplete, but it can also be biased in that parents of children with adverse outcomes can provide a different history of medication use than parents of children without adverse outcomes. So, one of the major reasons for using automated databases is this potential for obtaining a relatively complete unbiased history of medication use.

Another reason for using automated databases is that the medical encounter files provide the possibility of efficiently detecting adverse fetal outcomes. For example, if we are interested in pyloric stenosis, the medical records of children in the first year of life can provide an excellent way of detecting pyloric

stenosis. Other examples of studies done in Tennessee using automated databases include those related to metronidazole. This drug is now very frequently used in pregnancy, but initially there was a good bit of concern because metronidazole is both teratogenic and carcinogenic in several animal species. Tennessee Medicaid data were useful in showing that this was not the case in humans. These data were also used to document the occurrence of adverse pregnancy outcomes related to the use of angiotensin converting enzyme (ACE) inhibitors.

There are some difficult issues related to the use of automated databases as well. In drug epidemiology, there is often great initial enthusiasm about the concept of using databases. There are computerized files for a defined population that contain all the drug use information and all the medical outcomes, everything that is needed. But there are also a lot of details that must be addressed, and often the initial enthusiasm abates. Complexities specific to the area of fetal drug effects involve two factors. The first is that it requires studying two people—the mother who is taking the drug that leads to the fetal effect, and the child who experiences the adverse outcome. And so, the records of two people must be linked together. The second issue has to do with the window of time during which medication exposure can cause harm, which is very specific. It extends from somewhere around the last menstrual period through the date of birth. Unless dealt with carefully, the issues around identifying the study population and the timing of the window of interest can lead to major flaws in study design. The other issues involved are generally similar to those in other areas of drug epidemiology.

The first challenge is to accurately define the study population: who are the children and who are their mothers. Next, for each study subject, when was the child born and what was the date of the mother's last menstrual period. While these are relatively straightforward questions, they must be handled with some care. An additional challenge is to accurately define the exposure. In these databases, a woman who is not enrolled in the system must be distinguished from a woman who does not receive the drug of interest. In the Medicaid population in particular, women often enroll in order to receive prenatal care. And so, we must be very certain that a woman who enters Medicaid during the third trimester is not classified in the study as someone who was unexposed to a potential teratogen in the first trimester.

There are also issues related to detecting adverse effects. One of the well-known problems in using vital records for this purpose is the incompleteness of birth defect information on birth certificates. And then, of course, there are issues related to potential confounding in any study of medication effects, particularly when the drug is used to treat a specific condition. Is it the disease that leads to the increased risk or the drug itself?

Following are examples of how we can get this information on a population from automated databases. In an HMO database there are records of women who are in the HMO and records of children who are in the HMO. The first step a researcher might consider is to link the records of mothers and children to identify exposed pregnancies and look at the children's outcomes. The problem is that such linkages typically are not necessarily very good. Some organizations have family identification numbers that can provide a good method of linking mothers and children. But most Medicaid programs, for example, do not. If the child's last name is different from the mother's, how can they be linked to each other? One thing that has been done is to link the HMO records with vital records, which typically have both the mother's and the child's name in a single place. We then have a good idea of the answer to the questions: Who were the children and who were their mothers?

Another question is how to identify the date of birth of the child. With HMO records, we might consider going from the enrollment record of the mother to her birth hospitalization record. Unfortunately, there are a couple of problems with that. In many organizations, 2% to 5% of hospitalization records are missing. In Medicare data, 2% to 5% simply get lost in the transit from the carriers or the fiscal intermediaries up to the Health Care Financing Administration (HCFA). In Medicaid data, about 5% of women with births do not have a Medicaid hospitalization record for one reason or another. Conversely, birth certificates are very complete and provide a very good answer to the question of when was the child born. And the last ques-

tion is, when did the pregnancy begin, or what was the date of the mother's last menstrual period? Some studies have been conducted using the date 9 months before the date of birth, but this is not an accurate assumption answer because many children are born with shorter gestational periods. This could result in the inclusion of exposures that actually occurred as much as 2 months before the last menstrual period.

So, for all of these reasons, linkage with vital records has been a kind of cornerstone of all the work done with Tennessee Medicaid data. Vital records also can provide useful information about additional factors, such as birth weight or smoking during pregnancy for example, which can be important in analyses. But, of course, one of the real issues is that of confidentiality. At Vanderbilt, we have worked in partnership with the Tennessee Department of Health, and the use of vital records information has had clear public health purposes. An example is the greater effort to avoid use of ACE inhibitors early in pregnancy. But, nevertheless, there is some discomfort with linking files in a way that puts together information people did not anticipate would be put together. How this will be sorted out remains to be seen, but it continues to be an issue for this type of research.

In conclusion, automated databases have an important role in the study of fetal drug effects and, in particular, they provide an excellent and potentially unbiased source of information on drug exposures. They are most effective when linked with vital records, but this has the corresponding limitation of the issues of confidentiality and privacy. If some of these issues can be addressed, the trend of the future will likely be toward studies from multiple databases that increase the power to assess relatively rare but potentially serious birth defects.

Related Literature

Ray WA, Griffin MR. The use of Medicaid data for pharmacoepidemiology. Am J Epidemiol 1989;129:837–49.

Piper JM, Ray WA, Griffin MR, Fought R, Daugherty JR, Mitchel E Jr. Methodologic issues in evaluating expanded Medicaid coverage for pregnant women. Am J Epidemiol 1990;132:561–71.

de Jong PC, Huijsmans AA, Nienhuis HE, Nijdam WS, Zielhuis GA, Eskes TK. Validation of a questionnaire on medical drug use during pregnancy. Am J Epidemiol 1991;134:998–1002..

Piper J.M, Mitchel EF, Jr., Snowden M, Hall C, Adams M, Taylor P. Validation of 1989 Tennessee birth certificates using maternal and newborn hospital records. Am J Epidemiol 1993;137:758–68.

Piper JM, Ray WA, Rosa FW. Pregnancy outcome following exposure to angiotensin-converting enzyme inhibitors. Obstet Gynecol 1992;80:429–32.

Piper JM, Mitchel EF, Ray WA. Prenatal use of metronidazole and birth defects: no association. Obstet Gynecol 1993;82:348–52.

Thapa PB, Whitlock JA, Brockman KG, Gideon P, Mitchel EF, Roberson P, et al. Prenatal exposure to metronidazole and risk of childhood cancer. Cancer 1998;83:1461–8.

Discussion Sessions

Following is a synopsis of the major points of the discussions within the five groups that took place over the course of the 2-day workshop. It is intended to present the opinions and advice offered by those who attended, and to promote an understanding of the issues involved in actively monitoring the effects of medication use during pregnancy. A consensus of opinions was not sought from the discussions and was not obtained.

Discussion Session #1:

Which drugs are important to monitor through post-marketing surveillance?

Which drugs are realistic to monitor through post-marketing surveillance?

Deciding which drugs are important to monitor is both an easy and a difficult task. The simple answer is that it is important to monitor all drugs. Any substance that has a biological effect has at least a theoretical potential to impact the fetus when taken by the mother. New drugs are developed specifically because they have effects different from those of existing drugs, and definitive information about their effects on human pregnancy is not available at the time they are marketed. Among drugs currently on the market, there is insufficient information to accurately define the safety margins of most, and even those within a given class can vary widely in their teratogenic potential. Therefore, all drugs should be eligible for monitoring. While the ideal is to collect information about all pregnancy exposures and outcomes, the reality is that, at best, there are imperfect methods and finite resources available to do this. Thus, the question becomes not which drugs are important but which is it realistic to monitor. The need to prioritize is unavoidable, despite the fact that all exposures are important.

The concept of which drugs are important and which ones it is realistic to monitor can be viewed from two sometimes conflicting perspectives. The first is that of the individual, "What does this mean for me and my child?" This perspective concerns the consequences of a specific exposure or a specific treatment decision during pregnancy on an individual mother and child. Weighing the relative risks and benefits of a specific medication to an individual patient and her pregnancy is the situation faced every day by women, health care providers, and counselors in clinical practice. The second perspective is that of the population, "What is the increased burden to society associated with exposure to this drug in pregnancy?" This perspective concerns the frequency with which pregnancy exposures occur and how to generate information that will help prevent unwanted outcomes that place the greatest clinical and financial burdens on society. This perspective, whether implicit or explicit, is the situation faced by researchers, public health workers, and government officials when making decisions about the application of resources for post-marketing surveillance.

There are a number of possible ways to prioritize drugs for surveillance. All drugs could be prioritized, or all new drugs could be monitored and those already on the market prioritized. One approach for drugs already on the market might be to apply statistical and other data-mining techniques to existing information from health maintenance organizations, adverse event reports, and other sources to assess, on a population basis, which drugs are of concern. Another approach, applicable to both old and new drugs, might be to rank each drug according to a predefined weighting of the factors considered important. A threshold would be set above which monitoring of a drug would be conducted, taking into account the various characteristics of different drugs and the coexistence of more than one weighted factor. This ranking could also be used to assign different levels or intensity of monitoring to different drugs based on combinations of the weighted factors.

One important factor to consider when prioritizing drugs is their frequency of use by women of child-bearing potential. The greatest ethical obligation might be to give first priority to the drugs that are used by the largest number of reproductive age women. This approach assumes that the most frequently used drugs will be those to which the

largest number of pregnancies are inadvertently exposed. Prioritization could be by the absolute frequency of use, that is, the drug is monitored if its usage exceeds a certain level, or by the relative frequency of use compared with all other drugs, that is, the top 10 most frequently used drugs. Related considerations include the frequency and duration of treatment likely to be received by an individual.

Another factor to consider when prioritizing drugs is the extent and quality of information known about the teratogenic effects of the drug. Drugs given high priority in this scheme might be (1) those for which premarketing animal studies show a teratogenic effect; (2) those with specific pharmacologic or pharmacokinetic properties that make them of concern, for example, drugs that influence growth; and (3) drugs for which case reports or clinical trials indicate a possible teratogenic effect. Priority might also be given to drugs that represent new molecular entities or new classes of drugs that have not previously been used or that have a unique mechanism of action. However, when considering these factors it is important to realize that the effect of a drug in human pregnancy cannot necessarily be predicted from animal studies or from its structure, class, or similarity to other compounds. In addition, while the primary goal of postmarketing surveillance is to identify previously unrecognized adverse effects of medication use in pregnancy, it might also be a priority to monitor drugs that are known teratogens and those specifically labeled not for use in pregnancy. These drugs might receive priority in order to assess whether ongoing exposures occur, identify circumstances under which exposure can be prevented, and characterize the full extent of the risk.

The condition for which the drug is used is another factor to consider. Of particular interest here might be drugs used during assisted reproductive procedures because of their intentional use at the time of fertilization and during the earliest stages of development. One could also consider that drugs used to treat life-threatening conditions, those used during surgery or other hospitalization, and those given for pain management be fully characterized. Priority might be given to drugs used to treat chronic illnesses common in women of childbearing age, such as asthma or epilepsy; conditions unique to pregnancy such as hyperemesis; behaviors known to be risky during pregnancy such as smoking; and acute conditions not related to pregnancy that can arise. It might be essential to identify which drugs among those used for a particular condition carry the least risk to the fetus so that considered decisions about treatment that best protect the health of both mother and fetus can be made. This could require monitoring all drugs used to treat a single condition in order to evaluate which is the safest alternative or to compare the effects of a new medication with those of older more commonly used drugs.

It is also important, when establishing priorities, to consider the settings in which the drug might actually be used. Medications may be prescribed or used for conditions other than those for which they were approved, so-called off-label use. There are also circumstances in which more than one drug is used simultaneously, either to treat a single condition such as epilepsy or to treat coexisting conditions. It must be kept in mind that drugs taken simultaneously can interact to produce effects in ways not predicted from their individual use. An additional factor is the level of concern or anxiety about the drug experienced by health care providers and the public. Even if a drug does not meet other weighted criteria, its monitoring might be considered a priority in order to generate information to clarify its risk or the absence of risk in order to alleviate concerns and facilitate the use of effective medications when needed.

While the ideal situation might be to monitor all drugs, there are a few categories that might be considered of lower priority. These could include drugs used exclusively by men, those used only topically, and those with a long history of frequent use without reported adverse pregnancy outcomes. One would expect the level of embryonic or fetal risk under these conditions to be low. In addition, while the focus of this discussion is on prescription medications, similar information is desirable for other biologically active compounds, including over-the-counter drugs, nutritional supplements, herbal preparations, and vaccines. These compounds are widely available and widely used specifically because they have biological effects. Their use raises questions and concerns similar to those for prescription drugs, including the potential for interaction with other prescription and nonprescription agents. For

these reasons, it might be considered a higher priority to monitor certain of these preparations than to monitor certain other prescription drugs.

Which outcomes are important to monitor through post-marketing surveillance?

Which outcomes are realistic to monitor through post-marketing surveillance?

When considering which outcomes to monitor, there is tension, as with monitoring drugs, between what is important and what is realistic, between individual and societal perspectives. An individual drug might produce one or a spectrum of effects, some of which are related to the exact timing of use during gestation, the dose and frequency of use, metabolic changes during pregnancy, the genetic constitution of mother and fetus, the presence of associated conditions, and other factors. The simple answer is that it is important to monitor the full range of potential outcomes on the embryo, fetus, child, and adult. This range of outcomes is extensive and includes the immediate pregnancy result (stillbirth, prematurity, growth retardation, neonatal death, or multiple birth) and the presence of major structural defects, multiple defects and syndromes, minor structural defects, metabolic and physiologic abnormalities (for example, jaundice and respiratory distress), functional deficits (for example, mental retardation, learning disabilities, and hearing loss) and long-term outcomes (for example, carcinogenicity, infertility, and age at death). To evaluate all of these requires comprehensive assessment of the whole person, his physical, functional, developmental, emotional, and long-term health, relative to the exposure. This is not practical on a broad scale. The question then becomes how to prioritize the outcomes monitored to provide information that is timely, accurate, and useful in making decisions about pregnancy management.

When considering which outcomes to monitor, as with drugs, one approach is to rank them according to a predefined weighting of the factors considered important. Such factors could include the frequency with which the outcome occurs and its severity, both in terms of the impact on the child and family and the cost of treatment and care. A key concept here is that the outcomes monitored must be understandable and of relevance to patients and health care providers. Miscarriage, prematurity, and low birth weight are the most common adverse outcomes, each occurring in approximately 5% to 15% of recognized pregnancies. The acute and long-term care of many of these infants who survive can be costly. Among the structural defects, those that are severe, permanent, and associated with significant functional impairment or with high attendant costs, and those that represent dreaded events from the patient's perspective might be of highest priority. While approximately 3% of live births have a structural or chromosomal abnormality, few teratogens exhibit a simultaneous effect on multiple organs. Most produce specific defects or a specific range of defects that occur much more rarely. Those individual defects that occur most frequently in the population, such as polydactyly, are not necessarily the most serious or costly.

While adverse pregnancy outcomes and structural defects might be dreaded events, primary functional and developmental disabilities, such as mental retardation, neurobehavioral problems, hearing and vision loss, and learning disabilities, are important from both the patient and societal perspectives. These conditions affect a person's quality of life and ability to contribute to society. One challenge of prioritizing is to assess, for example, whether cleft palate deserves the same priority as mental retardation. Unfortunately, there is no reliable and widely accepted means of objectively quantifying the relative burden of these factors.

An alternative approach is to select the outcomes to be monitored for a specific drug based on the existing knowledge about that drug, its mechanism of action, the results of animal studies, and the class of drugs to which it belongs. In this scheme, the outcomes monitored would be different for different drugs. This might have the advantage of limiting the range of outcomes evaluated while maintaining a logical approach to the application of resources and efforts. The disadvantage, of course, is that drugs can have unforeseen effects that are not always predictable from information available at the time they are marketed. Indeed, it is the very occurrence of a new or unusual pattern of abnormal growth and development or an unusual combination of defects that might most accurately indicate a

new teratogen. In such a scenario, the more relevant focus would be on outcomes that are beyond the range of what is typically seen.

Regardless of which approach is used for prioritizing, it is important to consider the ease and practicality with which an outcome can be defined, identified, and monitored, and the cost of such monitoring. The ability to accurately define a case is a key concept in public health surveillance. While this might be relatively straightforward for immediate pregnancy outcomes and major structural defects, it can be far less clear-cut for learning disabilities, behavioral abnormalities, and even hearing impairment, which can be present to varying degrees. Similarly, the older the age at which a condition is recognized relative to the prenatal exposure, the more difficult it can be to systematically monitor and evaluate the relationship between them. Among the developmental disabilities, it might be most realistic to monitor severe mental retardation simply because it can be recognized more easily and comparatively earlier in life. Ascertainment of milder degrees of intellectual impairment and other conditions might be practical only if limited to a specific period of time or age after exposure. Long-term outcomes such as cancer, infertility, and impact on subsequent generations, while extremely important, are particularly difficult to monitor and would require identification of subjects for future longitudinal follow-up at considerable cost.

A related issue is the availability of data to evaluate specific outcomes. The accessibility of information about health outcomes for large numbers of people is inherently difficult because of concerns about data privacy, patient autonomy, and potential consequences of the use of such data. Learning disabilities and behavioral disorders are frequently evaluated and diagnosed through schools, the records of which are not part of the health care system. Widespread use of prenatal diagnostic testing has made it imperative to ascertain the occurrence of severe defects among elective terminations to accurately assess the impact of a drug exposure. Documentation of those defects and access to those records can be problematic for a number of reasons.

Which levels of risk are important to monitor through post-marketing surveillance?

Which levels of risk are realistic to monitor through post-marketing surveillance?

The goal of post-marketing surveillance is to generate the minimum level of knowledge needed to understand the potential risk of drug exposures during pregnancy and to communicate that risk to pregnant women and their families. While studies frequently measure the effects of drug exposure relative to that of a reference group or population without the exposure, patients often think in terms of absolute risk: "Am I going to have a problem?" In contrast, the societal perspective often focuses on the burden of illness in the population that results from an exposure, the attributable fraction. How these concepts are measured and communicated to patients and health care providers is of utmost importance.

The concepts of risk and safety are intertwined and inseparable. The term "safety" implies the absence of risk, which is impossible to demonstrate conclusively. There is always a confidence interval surrounding the estimate of an association, or lack of an association, between an exposure and an outcome in any study. Whether a drug can be considered "safe" depends to a large extent on the width of that confidence interval and the perspective of the person interpreting it. The interpretation can vary between societal and patient perspectives and with the circumstances of individual patients and families. For example, some might consider an upper 95% confidence limit of 1.5 around the relative risk of an exposure as indicating safety. However, it might be less clear whether a lower 95% confidence limit of 1.2 would indicate lack of safety. A more helpful concept, when communicating with patients and families, might be to talk about the margins of uncertainty around what is known about a drug's effects.

Setting priorities for the level of risk to monitor will vary depending on the specific drug and outcome under consideration and the factors used for prioritizing. These might include the indication for and benefit from use of the drug, the available information about the drug's mechanism of action and potential effects, the frequency of its use among

women of reproductive age, the ability to obtain accurate data about the exposure and outcome, and the level of concern about the drug or class of drugs. Clearly, the first priority for new drugs is to be certain they do not carry a teratogenic risk as potent as that of thalidomide. This is both important and realistic, as a relatively small exposure cohort would be required to detect risks of that magnitude even for individual structural defects.

The practical question thus becomes whether and when is it realistic to monitor for risks of lesser magnitude. The sample size required to measure a specific level of risk or lack of that risk is related to the frequency with which the drug is used, the frequency of the outcome, the specificity with which these can be defined, and the presence of associated factors. It seems reasonable that patients and clinicians would want to know if a drug increases the risk for a specific outcome by 50%, a relative risk of 1.5. If that is practical to assess, it is a reasonable goal. However, this might not always be practical from the standpoint of conducting studies. A general range of relative risks from 1.5 to 5 might be a more realistic goal for surveillance. Where within that range a specific study is able to target will depend on the factors discussed, as well as the cost and feasibility of conducting the study.

Detecting a relative risk of 1.5 to 3.0 might be realistic for reproductive outcomes and for all major structural defects together, as these outcomes occur in approximately 3% to 15% or more of recognized pregnancies. For serious developmental disabilities and other neurologic problems that require special education services, it might be possible to detect a twofold increase provided these diagnoses can be reliably ascertained. The more common individual structural defects that occur at baseline in the range of 1 in 1,000 pregnancies are also potentially practical to monitor, but can require larger studies. Detecting relative risks of 5 to 10 for these conditions might be more feasible. However, rare individual defects that occur at baseline in the range of 1 in 10,000 pregnancies are almost impossible to measure. Monitoring for any but the largest increases in these conditions is probably not practical, regardless of their severity or importance. Unfortunately, it can be difficult for individuals to place such rare risks into the context of their own experience when faced with an exposure. The perception of risk can appear inflated when there is a large degree of uncertainty about a drug's effects.

An alternative approach might be to set a critical absolute level above which to monitor for adverse effects. Rather than measuring the frequency of an outcome relative to an unexposed population, this approach would compare the frequency to a predefined level. Only increases above this level would be considered of concern. Increases in more rare conditions that did not reach this critical level would not be evaluated even if, in actuality, the drug increased their risk. However, while a single critical level of risk might be used for rare conditions such as individual structural malformations, different critical levels would probably be needed for conditions that occur more frequently.

In reality, the approach that would yield the most information might be to conduct different levels of surveillance for different drugs and outcomes. For example, it might be more realistic to exclude a high level of risk, but to accept uncertainty about more moderate risks, when evaluating drugs that are rarely used, outcomes that are extremely rare, or drugs that are of clear benefit in the treatment of serious maternal conditions for which there are no less risky alternatives.

Discussion Session #2

In this session, three separate models for conducting post-marketing surveillance for the fetal effects of medications in pregnancy were proposed. The five groups were asked to discuss the characteristics, strengths, and limitations of each. While based on studies from the literature, these theoretical models were constructed to illustrate different methodologic approaches and data sources that could be used.

Model A—Example of a Prospective Cohort Approach

Health care providers caring for pregnant women who have been exposed to a medication during pregnancy call a central 1-800 number to enroll their patients in the surveillance program.

Providers learn about the program from a variety of sources such as pharmaceutical company representatives, mailings, the Physicians' Desk Reference, and advertising in professional journals. It is expected that 50 to 300 women with exposure to an individual drug could be enrolled by this method each year. The actual number will vary for individual drugs depending on the frequency of use among women of childbearing age in general, among pregnant women in particular, among different subpopulations, and other factors.

The health care provider is mailed a registration form to fill out about the woman and her exposure. Minimum data asked about each woman enrolled include her race and ethnicity; age or birth date; date of her last menstrual period (LMP) and estimated date of delivery (EDD); gestational age at enrollment in the program; dates, dose, frequency, and duration of all medications taken from one month prior to conception through the date of registration, including vitamins and dietary supplements; information about potential confounders, including infections and use of tobacco and alcohol; information about prenatal testing including the type, date, and outcome of each prenatal test; and the physician's name and contact information.

Approximately one month after the estimated date of delivery, the health care provider is mailed a follow-up form to fill out about the outcome of the pregnancy. Minimum data asked about the outcome include the date of delivery; gestational age at delivery; sex; birth status (live birth, stillbirth, or termination); birth weight, length, and head circumference; presence of any physical birth defects and other medical conditions or abnormalities; type, date, and result of any test performed on the infant after delivery; and nature of any specialty medical care planned after discharge from the delivery hospital.

Frequencies of outcomes of interest among exposed women who were enrolled early in pregnancy before prenatal tests were performed are compared with baseline frequencies reported in the literature. Relative risks, confidence intervals, and study power are calculated for each comparison. The findings of the surveillance are tabulated yearly and mailed to health care providers who have participated in the surveillance and to other interested parties upon request. When findings, either positive or negative, of a predetermined statistical significance are reached, the results are published in a peer-reviewed journal.

Strengths: Prospective cohort studies are particularly efficient for evaluating specific rare exposures because they begin by selecting pregnancies with the exposure of interest. They allow calculation of the risk of an outcome among those exposed. This is probably their most important strength. If exposed pregnancies are ascertained from individual health care providers, as in the model described previously, they can be particularly useful for generating early signals for new drugs as they appear on the market and begin to be widely used, and for drugs about which there is *a priori* concern regarding their potential effects. The prospective nature of this approach provides the opportunity to evaluate a range of effects of individual drugs on the fetus and newborn and to more fully characterize their impact.

Because information about medication use is obtained directly from the woman's health care provider while she is still pregnant, the exposure data can be less subject to recall bias than in retrospective approaches. Similarly, because women are enrolled in the surveillance program before the outcome of the pregnancy or the presence of fetal conditions is known, the data are free from the selection bias toward more severe outcomes that can be present among adverse event reports.

This approach to collecting prospective reports of exposures in pregnancy has been used in a limited manner by some pharmaceutical companies and other investigators to monitor the effects of certain drug exposures during pregnancy. In its simplest form, a cohort of 100 or so pregnancies exposed to a single drug would be relatively inexpensive to ascertain. Similar evaluation of therapeutic classes or groups of drugs has been conducted using shared resources. Examples of registries using this shared approach include the Antiretroviral Pregnancy Registry and the Antiepileptic Drug Pregnancy Registry.

Limitations: For the very reasons that the prospective cohort approach is efficient for evaluating rare exposures, it is less efficient for evaluating

rare outcomes. It is best suited for the combination of a rare exposure that increases the risk of a relatively common outcome. It is a strong approach for identifying potent teratogens such as thalidomide or Accutane®. However, the cost of recruiting a sample large enough to identify drugs that produce moderate increases in less common outcomes or to ensure that such an increase is not present would be prohibitive.

Because exposed pregnancies are ascertained through voluntary reports from health care providers, they can have characteristics different from and nonrepresentative of the general population. This selection bias could lead to inaccuracies in the conclusions drawn and is compounded by the fact that a comparable control group is not recruited. Comparison and interpretation of findings are made relative to published data from other populations that might not necessarily be comparable.

An important additional consideration is the source and quality of the outcome information obtained. In the model described previously, outcome information is obtained from the mother's health care provider, rather than the child's, in order to facilitate outcome reporting. In reality, many obstetricians have little contact with or knowledge about an infant beyond the delivery room. Outcome data obtained in this way can be of variable quality and might be nonexistent for internal defects, conditions recognized after the immediate newborn period, and developmental or functional outcomes. This is particularly true if the reporting provider is not an obstetrician but a subspecialist, such as a pulmonologist who manages the mother's chronic asthma. In this situation, there is the potential for false negative reports of outcomes and the lack of recognition of a signal when an adverse association might actually exist.

The prospective approach, as with any follow-up study, also has the potential for significant loss of subjects between the time of enrollment and the request for outcome information. This loss could result from the mother changing health care providers during pregnancy, lack of time or ability of providers to answer questions about the outcome, concerns about potential liability, and the like. In the model described previously, systematic consent from the pregnant woman is not necessarily obtained. All of these factors can limit the sample size and generalizability of the study results.

Model B—Example of a Case-Control Approach

Infants with the outcome of interest are ascertained from population-based state surveillance systems. Infants without the outcome of interest are randomly selected from hospital delivery records or birth certificates from the same population. Mothers of the infants are mailed information about the study as soon as possible after delivery, then contacted by telephone to determine their willingness to participate and to set a time to conduct a telephone interview. Informed consent is explained and permission forms for access to mother and child hospital records are mailed for signature.

An example of expected enrollment for all major birth defects would be 300 infants with defects and 100 infants without defects per year. Examples of expected enrollment for specific defects would be (1) common defects: 25 infants with cleft lip with or without cleft palate per year and 21 infants with a conotruncal heart defect per year; (2) moderately frequent defects: 7 infants with gastroschisis per year and 8 infants with craniosynostosis per year; and (3) rare defects: 4 infants with anotia or microtia per year and 1 infant with Ebstein anomaly per year. If data from more than one state system were combined, the sample sizes could be substantially increased.

Maternal interviews are conducted using an extensive computer-assisted telephone interview that takes approximately one hour to complete. To be included in the study the interview must be completed within 24 months of delivery. Data obtained from each mother include the mother's age, race and ethnicity, and education level; history of previous pregnancies (live births, stillbirths, terminations, birth defects, and fertility); the dose, frequency, duration, and timing in gestation (exact dates, month, or trimester) of all medications taken from one month prior to conception through the date of the interview, including vitamins and dietary supplements; information on maternal illnesses during pregnancy, both chronic and acute; information about other maternal exposures during pregnancy,

Table 1. Example of screening for associations of Drug X with different defects using the case-control approach (Total number of subjects = 16,261; exposure rate in first trimester = 65/16,261 = 0.4%)

Defect	Exposed cases	Exposed controls	Relative risk
Vascular disruption	5/1,237 = 0.40	60/15,024 = 0.40	1.0
Male genital defects	6/1,465 = 0.41	59/14,796 = 0.40	1.0
Inguinal hernia	7/1,240 = 0.56	58/15,022 = 0.39	1.5
Cleft lip + palate	9/832 = 1.1	56/15,439 = 0.36	3.0 ($p<0.01$)

including infections, fever, occupational exposures, nutrition, and use of tobacco, alcohol, caffeine, and illicit drugs; information about prenatal testing including the type, date, and outcome of each prenatal test.

Outcome information is obtained from the surveillance system records of birth and pediatric referral hospitalizations for infants enrolled in the study. All information is reviewed by a clinician trained in birth defects, genetics, and dysmorphology. Information obtained includes the delivery date; gestational age at delivery; sex; birth status (live birth, stillbirth, or termination); birth weight, length, and head circumference; description from medical records of any physical birth defects, other medical conditions, or any other abnormalities noted; type, date, and result of any test performed on the infant after delivery; and nature of any specialty medical care planned after discharge from the delivery hospital. Odds ratios and confidence intervals are calculated for medication exposures and outcomes of interest. Results of case-control analyses are published as descriptive and analytic manuscripts in peer-reviewed journals.

Strengths: Retrospective case-control studies are particularly efficient for evaluating specific rare outcomes because they begin by selecting infants with and without the outcome of interest. Because both exposure and outcome have occurred at the time the study is undertaken, this method is convenient and rapid. Studies assess the magnitude of exposure relative to that of an unaffected comparison group, and thus can evaluate negative as well as positive associations. For these reasons, case-control studies are particularly suited for evaluating individual structural defects and for generating hypotheses through rapid assessment of the strength of associations. However, they can also be used to evaluate any outcome that is clearly defined and identifiable, including syndromes diagnosed through physical examination by a dysmorphologist.

The model described is based on methods used by the National Birth Defects Prevention Study (NBDPS), which is supported by the Centers for Disease Control (CDC). Specific strengths of this model include the fact that pregnancies with and without the outcome are ascertained from the same population-based surveillance system, which can help minimize selection bias between these two groups. The fact that outcomes are ascertained in a uniform manner through documentation with medical records and review by trained clinicians provides particularly high-quality outcome data. The inclusion of pregnancies electively terminated after prenatal diagnosis increases the sample size available for study and helps minimize any bias related to the availability or use of prenatal diagnosis and elective termination. Similarly, using a standard questionnaire to conduct direct interviews of women who have had both affected and unaffected pregnancies allows ascertainment of the actual pattern and timing of medication use, which can help reduce exposure misclassification. Maternal interviews also provide the opportunity for assessing potential confounders and for evaluating multiple simultaneous exposures.

Limitations: For the very reasons that case-control studies are efficient for evaluating rare outcomes, they are less efficient for evaluating rare exposures. The approach is best suited for the combination of a relatively common exposure that increases the risk of a rare outcome. Because

women of reproductive age are generally healthy, many individual prescription drugs are used relatively rarely in pregnancy on a population basis, in the range of perhaps in one in a thousand pregnancies. In these instances, the sample size required to evaluate any but very large increases in outcomes with a case-control approach can be prohibitive.

Typically, case-control studies are designed to evaluate specific outcome-exposure associations. This requires an *a priori* decision about which outcomes should be evaluated before a study is undertaken. Such designation precludes the opportunity to monitor for unusual or unanticipated effects or for a spectrum of effects from medication use. In the model outlined, birth defects registries can be an effective source of ascertainment for structural malformations, but might not include less obvious manifestations. Ascertainment of other outcomes, such as miscarriage or autism, for example, would require additional efforts even if a retrospective case-control approach were used. In addition, unless interview questions are sufficiently open ended to include new drugs as they are marketed, the approach will have limited value to monitor new exposures over time.

While the inclusion of pregnancies electively terminated after prenatal diagnosis of defects is a strength of the model, confirmation and characterization of anomalies among electively terminated pregnancies is often less reliable than among live born infants. In addition, some internal malformations, such as cardiac defects, might not be evident or identified until weeks or months after birth. These can be underrepresented if newborn records are the primary source of case ascertainment for the birth defects systems, leading to inaccurate conclusions about a drug's effects.

While the quality of exposure information obtained directly from maternal interview is a strength of this method, it can also be problematic. Because information about pregnancy exposure is obtained retrospectively, often months after the pregnancy outcome, its accuracy depends on the mother's recall and can be subject to bias. A woman who has experienced an adverse outcome might remember her exposures during pregnancy differently from a woman with a normal outcome. Depending on how accurately drug use in early pregnancy is remembered, including changes in dose or frequency and when these changes occurred, there might also be significant exposure misclassification that could result in either underestimation or inflation of association with an outcome.

Participation in case-control studies is usually voluntary and depends on patient consent and cooperation. If there is a systematic pattern to participation or lack of participation, study results can be biased. For example, mothers of affected infants who had an exposure might be more likely to participate in a study because they had the exposure than mothers of unaffected infants who had the same exposure. In contrast, the length of time required to conduct a detailed interview might inhibit some subjects from participating, perhaps more commonly those who have had normal outcomes.

In the model outlined, findings from case-control studies are disseminated through publication in peer-reviewed journals. In general, studies that show a positive association between an outcome and exposure are published far more readily than those that do not show an association. In generating information about the effects of medications in pregnancy, it is important to also disseminate the results of studies that have sufficient statistical power to show the lack of an association and to interpret these in ways that are useful for patients and health care providers. Publication in peer-reviewed journals might not be sufficient to accomplish the latter.

Model C—Example of the Use of Existing Large Data Sets

Pregnant women are identified from the administrative records of a health maintenance organization (HMO). Pregnancy is indicated by the presence of a code for a health care provider visit, procedure, or hospitalization related to pregnancy or delivery. These records are then linked with the state's birth and fetal death records to obtain the pregnancy outcome (live birth or stillbirth); the child's name; and gestational age at delivery or the date of the mother's last menstrual period (LMP),

or both. Prescriptions filled by these women during pregnancy are then identified through linkage with the HMO pharmacy records. Information collected about each prescription includes the name of the medication, the date the prescription was filled, the number of days' supply dispensed, and the prescribed dose and frequency.

Exposure during pregnancy is defined as having filled a prescription one month prior to the LMP or at any time during pregnancy for which taking the supply dispensed would overlap the date of conception or any other time during gestation. Examples of the frequency of exposures in the first 120 days of pregnancy that have been obtained from Medicaid data include 19 exposures to angiotensin converting enzyme inhibitors among 106,813 pregnant women (1 of 5,622) and 1,387 exposures to metronidazole during a 6-year period. Outcomes are identified from coded diagnoses in the HMO administrative records for the children of women who filled these prescriptions. Diagnoses for stillborn infants are obtained from among the diagnoses coded for the mother. Frequencies of outcomes among women who filled a prescription for the medication of interest are compared with those among women who filled a prescription for (1) medications known to be nonteratogenic, (2) medications taken only during the last trimester of pregnancy, and (3) all medications in the surveillance. Relative risks and confidence intervals are calculated for each comparison.

If an increased risk for an outcome is detected at a predetermined level of statistical significance, or if there is clinical concern about a particular exposure-outcome combination, a case control study is conducted in which the records of infants in the HMO who have the outcome of interest and a random sample of those who do not are abstracted, along with those of their mothers. Minimum data obtained include race and ethnicity; mother's age or birth date; history of previous pregnancies (live births, stillbirths, terminations, and birth defects); the dates, dose, frequency and duration of medications taken during the relevant time in gestation for the outcome of interest, including vitamins and dietary supplements; information about potential confounders such as infections and use of tobacco, alcohol, and marijuana use; birth status (live birth or stillbirth); delivery date or gestational age, or both; sex; birth weight, length, and head circumference; and the presence of physical birth defects, other medical conditions, and any other abnormalities. Odds ratios for exposure to the medication of interest are calculated with corresponding confidence intervals and study power.

The results of case control studies, whether positive or negative, are described in a newsletter to HMO members. Positive results of an increased risk of an outcome after exposure to a medication are published in the peer-reviewed literature.

Strengths: Some of the same epidemiologic approaches used in models A and B, with their strengths and weaknesses, can be applied to the data in model C. However, there are also characteristics unique to the use of existing large data sets. The ability to link data from different sources has the potential to increase the sample size and range of information available for monitoring. Linking data from HMOs, Medicaid, vital records, military systems, and practice-based records, for example, could greatly expand the uses of these data. The existence of standard information about large numbers of pregnancies could facilitate the timely and relatively cost-efficient identification and evaluation of signals of potential adverse effects. Unexposed pregnancies and those without the outcome of interest from the same population could be used for comparison in nested cohort or case-control studies as the need arises. This might enable monitoring and evaluation of both common and relatively rare exposures and outcomes, as well as the simultaneous evaluation of multiple exposures and multiple outcomes.

In addition, because the information in such databases is not limited to pregnancy, there could be the potential to evaluate long-term outcomes and conditions, such as developmental and neurobehavioral abnormalities, not apparent in the first weeks or months of life. Currently, a large number of federal programs in the United States serve children with special needs. Linking data from these programs could potentially expand the ability to monitor these conditions.

Another strength of this approach is the use of prescription data to identify drug exposures during pregnancy. Because the compilation of prescription records is unrelated to pregnancy outcome and does

not rely on recall by the mother, the exposure ascertainment should be unbiased. This could provide an objective means of estimating the timing of medication use in pregnancy, particularly that used in the management of chronic conditions. Similarly, the standard collection of information from all patients in a database about associated diagnoses, behavioral factors, family history, and other conditions might facilitate the objective evaluation of these factors during analyses.

Methods being developed to use existing data for health care research unrelated to pregnancy constitute a growing infrastructure that might also benefit activities for surveillance of medications in pregnancy. An example is CDC's Vaccine Safety Datalink, which employs HMO data to conduct nested studies, when indicated, to evaluate potential risks of immunizations. The Centers for Education and Research on Therapeutics (CERTs), funded by the Agency for Healthcare Research and Quality, might also provide a platform on which to build the broader capacity for monitoring the effects of medication use during pregnancy using existing large data sets.

Limitations: While there might be great potential in the use of existing large data sets, it is important to realize that there are limitations to their current status as research tools. One is the fact that these data sets were originally established to maintain billing, administrative, or other records for nonresearch purposes. They can lack the specific health information needed for surveillance and research. For example, diagnoses might be lumped together into a nonspecific coding scheme that does not provide sufficient detail for analysis. Dysmorphisms and syndromes that involve a subtle pattern of abnormalities might not be coded unless they require a specific intervention or surgery. Similarly, outcomes such as developmental abnormalities, attention deficit, or learning disabilities might not be reflected in health care data unless a specific prescription is needed. The records might not include information about defects among pregnancies that are electively terminated. Also, information about potential confounding factors, such as smoking or alcohol use, might not be routinely recorded in a standard way unless they are of sufficient severity to require intervention.

Obtaining a sufficient sample size to monitor individual drug exposures and the less common outcomes would likely require linking several existing databases together. In a situation in which these data sets were established by different entities or for different purposes, they might not contain the same data in a standard format with the same accuracy and specificity. Complex algorithms and considerable time and effort might be required to establish their linkage and to assess the validity and quality of the resulting data. Unless such linkages were performed on a routine basis, there might be delays in the availability of data for monitoring.

Another issue in using existing data is that the subjects might not be representative of the general population. Conclusions drawn from their data might not reflect the general experience. This is especially true when data from health insurance plans are used. Those who subscribe to a particular plan can differ from those who do not in systematic ways related to their general health and their demographic, lifestyle and socioeconomic factors. In addition, individuals not infrequently change health care plans as they change jobs or places of residence. The characteristics of those who change might differ from those who remain. It is estimated that as much as 40% of an HMO population fluctuates at any given time, leaving approximately 60% available for monitoring longitudinal outcomes.

While prescription records available from these data can provide unbiased ascertainment of medication exposure, they do not necessarily reflect actual use of the drug. Patients can delay taking a medication, miss doses, discontinue the drug, or take a different dose than prescribed, for example. They might also obtain some prescriptions from a pharmacy or provider outside the health plan whose data are being used. In addition, not all drugs are included on the formularies of every health care system, and some newly marketed drugs might not be added for several years, if ever.

Finally, although the information in these databases preexists, their use for purposes other than that for which they were originally collected could challenge data sharing and privacy policies. When multiple data sets are linked, concerns over data ownership, responsibility for the conduct of studies,

and interpretation of results can require special consideration.

Discussion Session #3:

What other approaches would yield improvements in post-marketing surveillance?

How could these approaches be combined or coordinated to ensure that the goals of post-marketing surveillance are achieved? What would be the next steps?

The ultimate goal of post-marketing surveillance is to effectively address priorities that are both important and realistic. There is no single best approach and no single study design that will address all questions. Effective post-marketing surveillance activities can be conducted in a number of different ways, but they must be sustainable in order to succeed regardless of the methods used. In this regard, it can be helpful to build on the experience of long-standing surveillance activities in other subject areas, such as infectious disease or immunization, in terms of what works and what does not.

Foremost among the considerations in planning any type of surveillance activity is its scope and timely coordination with research and public health efforts. It is important not only to be able to detect a signal of an adverse outcome, but to further evaluate and then act upon it. Having those different pieces in place can allow resources to be put together quickly in the event a potential adverse signal is identified.

The willingness of patients and health care providers to contribute to surveillance activities can be critical to their success. There can be tension between the reluctance of patients and providers to divulge private information and their disappointment when information about the effects of medications used during pregnancy is not available. One consideration is not to overly burden health care providers with requests to participate in multiple studies or to report the same pregnancy to multiple sources. It is important to balance their participation with a realistic assessment of the time and energy required, and how that fits with the clinical care of patients. In this regard, pregnant women themselves can be a valuable resource for obtaining accurate health care records and additional information. But it is important that the value of their participation in generating new information with which to better counsel future women be communicated effectively. It is also important that those who conduct surveillance activities appreciate the benefit of providing something back to those who participate.

In terms of sustainability, periodic assessment and revision of the priorities, methods, effectiveness, and efficiency of any surveillance activity will be needed. Ongoing monitoring of the effects of drug use during pregnancy can be a dynamic process as changes develop in the pharmaceutical industry and in health care. The impartial interpretation of data and accurate dissemination of results are crucial to ensure scientific and administrative objectivity.

With those issues in mind, the following suggestions about other approaches to post-marketing surveillance were made during the five group discussions. They are listed individually, in no specific order, and are not necessarily mutually inclusive or exclusive. A consensus of opinions was not sought from the discussions and was not obtained.

1. Develop and use multiple surveillance components that build on existing methods to maximize the availability and utility of data on the effects of medication use during pregnancy. Information could be gathered from all available sources to look for signals of potential adverse effects of drugs. In addition, multiple drug-specific data collection and evaluation mechanisms could be employed so that whichever worked most efficiently and provided the most accurate information about a particular drug would be used. This would include a minimum of three approaches:

Continuing and expanding the adverse event reporting system that currently exists through more systematic and regular examination of data to generate hypotheses and signal unusual outcomes or patterns of outcomes that deserve further attention.

Linking existing records, such as those from health maintenance organizations and Medicaid, to monitor the outcomes of exposed pregnancies to

identify associations that deserve further evaluation with a detailed study or that warrant dissemination of information about risk or safety back to physicians and patients. While all drugs could be monitored, this level of surveillance would probably be most useful for those drugs more commonly used and about which there is sufficient data to identify individual associations.

Conducting exposure surveillance through a prospective cohort. A single common telephone number, perhaps 1-800-PRGNANT, could be used to uniformly ascertain pregnancy exposures in order to simplify enrollment and reporting. This level of surveillance would be most useful for new drugs and for those of particular concern, whether old or new. It also might make it possible to generate information about nonprescription exposures such as over-the-counter medications, herbal preparations, and dietary supplements.

Exactly how the various surveillance components might be implemented and who would implement them might vary depending on the circumstances. Some pharmaceutical companies might run their own data collection using whatever method is most appropriate for an individual drug, while others might participate through a central resource. Some drugs might be monitored through academic centers and others through teratogen information services depending on which approach is most appropriate. A centralized mechanism for uniform data collection could have some advantages in terms of conducting national surveillance and considering the various exposures as a whole, if these were the goals.

2. Explore, expand, and more fully develop existing programs to conduct targeted studies to closely evaluate signals of potential exposure-outcome associations generated by surveillance activities. A hybrid approach in which existing programs are reviewed, modified, integrated, and tailored to address specific circumstances might be most efficient. Examples of existing activities that could be used include:

The CDC-funded Centers for Birth Defects Research and Prevention, which have infrastructure and methods in place for conducting population-based surveillance for birth defects and interviewing mothers of infants with birth defects for case-control studies. Their data could be used to more closely examine specific drug-outcome associations. However, while combined data from these centers might facilitate evaluation of rare defects, they have limited ability to evaluate low frequency exposures. In this instance, linked records from large existing databases might be used to obtain larger sample sizes for case-control studies of rarer exposures.

The Organization of Teratology Information Services provides information, counseling, and support through state or regional telephone services for physicians and pregnant women in making decisions about the treatment of conditions during pregnancy and the management of exposed pregnancies. Some of these services have the ability to conduct follow-up studies of pregnancy outcomes after particular exposures. In some of these studies, all children are evaluated by a dysmorphologist. In other studies, children up to 7 years of age undergo developmental or neurobehavioral evaluation. These studies can provide the ability to evaluate outcomes not captured by other methods.

3. Set up standard rules to be uniformly applied for making decisions about configuring studies, participating in surveillance activities, collecting data, and publishing and disseminating information. When concern is raised about the effect of a drug, a common rule could be followed to evaluate whether a problem really exists. Consistent application of standard rules and decisions might require coordination through a supporting infrastructure. This could include a common mechanism for enrolling patients in studies, using standard forms for collecting exposure and outcome data, and a single 1-800 number for reporting regardless of the drug being monitored.

One means of setting such standard rules might be through an advisory panel of experts who would critically review drugs for eligibility, develop weighting criteria for prioritizing, interpret findings, and decide what additional steps could be recommended and how to guide their implementation. Decisions could be reached through discussion and agreement, particularly when dealing with the issues of false positive and false negative results. Frequent public announcement of associations without adequate verification could result in undue panic. In

contrast, failure to release associations might be interpreted as covering up adverse findings. An advisory committee could repeatedly examine potential signals and associations to ensure appropriate recognition of a signal, appropriate public announcement of teratogenic risk, and credibility of findings. To be successful, such an advisory committee would need to maintain good partnerships with other groups participating in surveillance activities in order to readjust its findings as data accumulate and to communicate those adjustments when publicizing information.

4. The idea of an advisory board might be taken a step further in the concept of a foundation, a public-private partnership along the lines of the Joint Information Center for the Study of Food Safety. The latter has been successful in conducting and funding surveillance and studies on food safety. Such a foundation could be made up of a variety of participants, including government agencies, private industry, international researchers, patient advocacy groups, legal and ethical specialists, and professional societies. Its goals could be to set priorities; conduct studies; serve as a clearinghouse for data; ensure the quality of information generated; and provide education and advice to clinicians, interested professionals, and the public. The foundation might address some issues itself through active surveillance and nested studies, and work with outside groups to conduct additional needed studies through collaborative relationships, both financial and intellectual. Such a foundation could ensure the participation of all interested parties without undue influence of any one constituent's interests, and be seen as a place from which reliable unbiased information was available.

5. It is important to keep open for consideration the concept of a longitudinal cohort to monitor outcomes over time and those, such as developmental, neurologic, or behavioral disorders, that are not evident or are not diagnosed shortly after birth. Unfortunately, this approach is limited in that a cohort defined in time could address only medications in use at that time and only those used with sufficient frequency to be evaluated by the cohort's size. The National Children's Study currently being planned by the National Institutes of Health and other agencies might have potential in this regard. However, an ongoing prospective cohort that continuously enrolls new subjects might be able to monitor the effects of new drugs as they are marketed.

In addition, existing data that might be used to monitor long-term outcomes, such as those from school programs, could be explored and tapped, if feasible, to coordinate with or contribute to other surveillance activities. This would require close attention to confidentiality, data privacy, and issues of consent.

6. Coordinate and collaborate with international researchers in related activities such as the International Clearinghouse for Birth Defects Monitoring Systems, the European Network of Teratology Information Services, and existing cohorts in other countries. Because of differences in the ability to compile and track health information about individuals, some countries might be able to generate or evaluate signals and longitudinal outcomes not practical in others. While not all of the medications used in one country are marketed in another, coordinating efforts might prove beneficial in generating information about those common to both. The frequency of international travel, employment and residence today emphasizes the need for information about the effects of medications in pregnancy regardless of where drugs are marketed.

For example, Sweden has a system of organized medicine in which the first prenatal visit at week 10 to 12 is free of charge. At this visit, a midwife conducts a 30- to 60-minute interview during which the mother is asked about a number of pregnancy exposures, including smoking, alcohol intake, and medications taken during the first trimester. A single standard form with this information is completed on each pregnancy throughout the country. Since 1994, data about drug exposures in all but 1% to 2% of pregnancies have been computerized and linked with the records of the Swedish Medical Birth Registry. This allows the regular and rapid generation of reports of the total number of pregnancies exposed to individual drugs and the total number of birth defects among those exposed. Because Sweden has 90,000 to 100,000 new pregnancies each year, there is now exposure information on approximately half a million pregnancies in the database. By contrast, the United States has approximately 4 million births per year, but no

quantitative system of exposure monitoring. Collaboration between these systems might use the Swedish data to conduct screening and set priorities for more detailed studies to be conducted in the United States.

Similarly, in the United Kingdom, office-based medical records are computerized throughout an individual's life and linked to maternal records. General practitioners record health information, including laboratory results, growth measurements, pregnancy exposures, and prescriptions written. Because general practitioners are the gatekeepers for social services, they also record information about social issues and school failure. This could make it possible to search backward in time for exposures that occurred prior to the onset of symptoms and to correlate with long-term functional disabilities. Such long-term correlations might provide information about medication effects that are difficult to assess in the United States.

7. Regardless of the type of activities developed or the reporting mechanisms used, a concerted effort is needed to develop public support for improving post-marketing surveillance for the effects of medications in pregnancy. There is a need to educate the public about why the lack of this information is an important public health issue. Without such support, improvements in the conduct and coordination of surveillance activities cannot be fully effective no matter how well conceived or executed.

Thoughts on an Organized Approach to Post-Marketing Surveillance for the Future

Elizabeth Conover, MS
Certified Genetics Counselor
Nebraska Teratogen Information Service

I represent the clinicians here and I would like to bring us back to why we met and why we have been talking about a better surveillance system. I have been running a Teratogen Information Service (TIS) for about 15 years and I think my experience is probably similar to that of a lot of other clinicians. Because I work primarily with a TIS, I say to someone every single day, "That is a really good question. Let me share what data we have. But I have to be honest with you, the data are not adequate to answer such a good question." I make apologies. My patients say, "Well, we got to the moon. We can cure leukemia. Why don't we know what this drug does to my baby? Why don't we know what drug is the appropriate treatment for my migraines while I'm pregnant?" Those are very reasonable questions, and I am embarrassed to say that we don't know the answers. It is not because we don't try. All teratogen specialists know how to get information. We know all the places to call, we have contacts everywhere. But the truth is, the information that exists is not very good. As someone said in our small group session, "Patients and clinicians would be shocked to know what poor data we use to make such incredibly important decisions about our children." I have always thought that children are important. But when I had two of my own it became very clear to me that, if I didn't have to take a risk, I would not take even 1/100th of 1% of a risk that something would interfere with my children growing up happy and healthy.

That is what every couple, every man and every woman, that comes into my office tells me. I run a project where we counsel very sick patients. They have cancer, or have had a liver transplant, or have bipolar disorder. They have something very serious and they are scared about the risks of pregnancy. Some of them are not yet pregnant but desperately desire to have children. They want reassurance that they can continue treatment of their serious medical condition during pregnancy with low risk. They want it to be safe, and I tell them, "We don't use the word *safe*." But, the goal is to keep the risk as low as possible. These people are already dealing with serious medical conditions and they deserve to have those conditions treated adequately during pregnancy. The fact is that, in many cases, we do not have the information that would allow us to do that.

So what do we do? I don't want to say that we don't answer patients' questions, because we do. We work with the trimester of pregnancy; we look for dose effects; we try to find older medications that have been better studied. But, in many cases, that isn't adequate for an individual woman. I didn't have to think very hard to provide you with some examples. On Monday of this week I saw a woman who has multiple sclerosis (MS). She has been taking interferon since her first event when she had optic neuritis and lost part of her vision. She was trying to become pregnant at the time, and so was not only dealing with the life-changing diagnosis of MS, but also the issue of whether to attempt a pregnancy. Well, at about that time she conceived and the question was whether to continue the interferon. Her neurologist said, "Well, we can stop the drug." In fact, that was his preference because he was concerned about the possible liability. In terms of treating her MS, however, he would like to continue the drug. He said, "Optimally, we think you will have a better long-term prognosis if you continue to take it." Now, anyone who deals with interferon knows that there is not good animal data about its effects in pregnancy, and there are only a handful of exposed human pregnancies reported in the literature. You can call the manufacturer to find out what information they have, but there are not very good data on the outcome of pregnancies exposed to interferon.

That same day, I saw a patient who had had a liver transplant and is now pregnant. She had not intended to become pregnant, but was unable to take oral contraceptives because of her medical condition and had failed barrier contraception. She did not believe in abortion, as is true with many

patients. She happened to be on a new immunosuppressant about which there are some questions from animal models about the risk in pregnancy. There is no information about the risk to human pregnancy in the literature, so I called the manufacturer. They knew of two exposed pregnancies, one of which had been lost to follow-up. The other ended in a spontaneous loss. So, there are essentially no data available for assessing the risk in this patient. I talked with her transplant surgeon and asked, "How badly does she need this? She is about 9 months out from her liver transplant." And he said, "Well, we can change her to another drug if that is what is best for the fetus, but it may not be the best thing for her." She could endanger her transplant and perhaps her life. And the transplant surgeon said, "So why aren't there any data?" Again, I said, "I am really just so sorry." I feel I'm groveling every time I say this.

On Tuesday, I came into the office and my secretary said, "You really need to take this call right away. This lady has been crying for 30 minutes and something is really up." So I talked to her and she said, "I am anxious, I am depressed, I have obsessive compulsive disorder. I have been taking Effexor® for this. I found out last week that I am pregnant and so I discontinued it. I think I am withdrawing. My husband says he is going to leave me because I am acting crazy all the time. I went to my doctor and he said, 'I don't think we know much about this drug in pregnancy. You can't take it and, I'm not sure, but you may have harmed the baby.' And then I cried for 3 days." And so I said, "Slow down and let's talk about it. There are always things we can do to help." I looked at the available data about Effexor®. There are a small amount of animal data that do not look too bad, so I reassured her about that. There were about 10 cases in some of the early clinical trials and then about 45 cases from some European data. Thank goodness for the European data. That was all that was available. And she said, "This drug has been used for a long time. Why don't we know more? It's the only thing that really works for me. I would never do anything to harm my baby, but my husband is having a fit. I can't take this any more. I'm thinking of hurting myself for doing this." This really should not be necessary. We should know the effects of Effexor® on the fetus. It has been out long enough and there have been plenty of pregnancies, but we haven't captured the data. The real question is, "Why don't we know?"

I belong to the Organization of Teratology Information Systems (OTIS) and we originally got together to make sure that we all had enough information to be able to counsel patients. We were really concerned about telling them the right things. After a couple of years, we realized that we knew what information there was, but that wasn't good enough. We were really concerned that, day after day after day, we were sitting on the hot seat and saying, "That is a really good question. We don't have enough data to assure you that it is safe to continue this drug during pregnancy. We can estimate the risk at between 5% and 20%" for example. And the patient would say, "Could you get a little closer?" That concern led us to do some of our initial studies. And OTIS has tried to begin some collaborative projects to try to get better answers because it is so important. There are a lot of days that I do not want to have to apologize one more time or hear from another desperate woman who wants me to come up with a way for her to get safely through her pregnancy.

And so I want to say how desperately I want us to come up with, not just one, but several different ways to go about gathering data to counsel women about the use of medications during pregnancy. I think that it will save money and that it is the compassionate way to approach this. Prevention works and it saves money. I think if we had even 1% of the money that went into treating children who have problems from drug exposures in pregnancy, we'd have much better answers for patients. The most important point is not to lose track, not to make this an academic issue. There are patients everywhere who rely on us to find a better way.

At OTIS, I think we have been very critical of our own data. We really want to generate reliable data, and initially we really were paralyzed by that. We wanted to design the perfect system. But, after awhile, we just started doing it and it became clear that we were going to learn something by doing it, and that each project has gotten better. So, I encourage us not to just continue discussing it, but to get going. It is important not to proceed thoughtlessly or without planning. But we need to get off the center and make the statement, "We

need this, we need money for it, let's go ahead and do some things." Most anything we do will be better than what we have already. My message is not to be paralyzed, to do something about it.

I'd also like to put in a plea for good data, so that I don't have to say to a patient, "Here is the good news. We don't think this drug causes birth defects. But the bad news is that we have no idea what it does to your baby's brain or immune system." And the patients say, "What kind of reassurance is that?" And of course it is incomplete, and that is not very good. It may be hard to gather data on outcomes that are not concrete or discretely measurable, but they are really important to patients. Children who are affected with fetal alcohol syndrome are not so harmed by the fact that they are a little small or even that they have a birth defect. What really handicaps their ability to live a normal life are the severe behavioral and intellectual problems. So I am making a plea that, even though we need to get going on this, we also need to know enough about all of the important end points to be able to suggest, with confidence, that it is okay for someone to take a particular drug during pregnancy.

And finally, I would like to mention that it is not sufficient to simply generate data. We need to think about how to disseminate the information, how patients will access that information, what is the appropriate presentation of the data, how best to describe risk, how to get the information to health care providers. That is a main thrust of what OTIS does, but I think we need to think of other ways as well. It could be with a pregnancy label for the drug, continuing education or other means. You can have all the great data in the world and it wouldn't make any difference in these women's lives if they don't have access to it. And so having data alone isn't sufficient. I care deeply about this and I am really hoping that something comes out of this meeting that will help my patients.

Thoughts on an Organized Approach to Post-Marketing Surveillance for the Future

Allen Mitchell, MD
Director, Slone Epidemiology Unit
Boston University Schools of Public Health and Medicine

I think it is important to both recognize that there are a variety of ways to approach these issues, and to try to reconcile the ideal with the feasible. The ideal is that we want to identify all increased risks for all drugs related to all specific adverse pregnancy outcomes. We want to know everything so that women who are anticipating a pregnancy can be fully informed. We can learn some things today or tomorrow, it might take quite awhile before we can learn other things, and there are things about the effects of drugs in pregnancy that we will never learn. What can be identified clearly depends on the frequency of a drug's use and the magnitude of the risk. The more common the drug, the more we can learn about it. The larger the risk, the more likely we are to identify it.

So, with that in mind, how should we go about thinking about this situation? There are really three overall objectives that are important and feasible to accomplish. The first is to identify major new teratogens quickly. There has been remarkable consensus at this meeting that this is the public health, research, clinical, and consumer goal of first priority. If resources were so limited that we had to choose a single priority, this should be the one. Also, to identify major teratogens quickly does not require a lot of attention to bias or confounding. The most prominent examples are thalidomide and isotretinoin, or Accutane®.

The other two objectives are more complicated from a research, scientific, and epidemiologic perspective because they get into often controversial areas of bias, confounding, and sample size. The second objective seems to me to be to identify teratogens with lesser risks. The association of valproic acid with neural tube defects is an example. We know there are a number of drugs that increase the risk of specific defects three- or fourfold, and there

are certainly debates about how many others go undetected.

The third objective is to estimate the ranges of safety or risk for all the remaining drugs. Doing so is both a statistical and an epidemiologic undertaking and the latter might be considered the more important perspective because it is not enough just to have numbers and "significance"; the estimated range of risk should be based on solid epidemiologic understanding so we can be reasonably confident that what we are seeing is correct. Consider a simple example: acetaminophen is used by about two-thirds of pregnant women. There are three or four publications that say something about the risk of this drug in pregnancy. We do not think it is a major teratogen because we have not seen a huge secular increase in defects with its use. We might identify, through a study, that acetaminophen does not appear to increase the risk of cleft palate, but we still have a wide confidence interval around that estimate of the risk. However, as time goes on and we collect more data, we can tighten that confidence interval. We will still have uncertainty about the risk of cleft palate with acetaminophen, but we can reduce the level of that uncertainty. The reality is that we will never reduce the uncertainty to the level of no risk. We can ultimately get to the point where, for a commonly used drug, we can say to a pregnant woman, "We can't prove that it causes no increased risk, but we can tell you that the biggest risk that it might cause is X." And that will give her something to go on.

With those three criteria in mind, I'd like to focus on the construct, the conceptual framework, for a monitoring system that we might be able to accomplish in the next 10 years. What I've done is break the objectives into two phases: how we might go about identifying major teratogens, and how we might go about identifying and reducing the uncertainty around drugs with lesser risks. In phase 1, in my view, we are talking about cohort studies that could involve pregnancy registries or databases. That doesn't mean that alert clinicians have no role. They should be encouraged to report clusters and observations, because virtually all of the teratogens known today were first identified by alert clinicians. But reliance on the alert clinician is an informal and unsystematic approach for monitoring drugs. If we want to identify major teratogens quickly, cohort studies such as those conducted by teratogen information services offer a lot of promise as a foundation on which to build. They have direct contact with pregnant women who have been exposed. A cohort of 100, or maybe 200, exposed women is sufficient to identify major teratogens such as thalidomide or Accutane®. The risk of Accutane® was identified within the first 36 prospectively followed pregnancies. Major teratogens such as these show themselves in small numbers of exposed pregnancies, and they can be identified very quickly.

So, if we follow a prospective cohort of 100 exposed pregnant women, we will get one of three results concerning major teratogenic effects: (1) there clearly is an effect, (2) there is no evidence of an effect, or (3) there is uncertainty about whether there is an effect. The point here is that if we see only 2 major malformations out of the first 100 pregnancies, we can be reasonably certain that the drug is not a teratogen of the magnitude of thalidomide or Accutane®. In the situation where we see weak evidence of an increased risk, we might conduct another cohort of 100 or continue the existing cohort to get a larger sample to be sure there is not a major teratogenic effect. But, for the most part, we will learn from a small cohort whether a drug is a major teratogen or it is not.

There are several strengths of this approach. First, it is not limited to evaluating only prescription drugs. We can identify any of a wide range of exposures to be followed. If we are interested in whether ginseng, for example, is a major cause of birth defects, we could conduct a cohort study of 100 pregnant women exposed to ginseng. If we are concerned about an over-the-counter drug such as pseudoephedrine, for example, we could recruit a cohort exposed to that drug. Second, there is no need to specify ahead of time the outcomes of concern. The cohort will identify them if there is a major increase. Third, bias and confounding are rarely substantive concerns in this setting because of the magnitude of the effects being considered.

In terms of the mechanism by which this might be set up, my view is that a centralized approach has advantages. Existing databases are a very valuable adjunct to these small cohorts. However, there are often delays between the time a drug is marketed and the time its exposure appears in these databases.

And, unlike small cohort studies or those of the teratogen information services, the use of databases does not provide the opportunity to collect additional information on factors such as smoking, alcohol, diet, and use of over-the-counter (OTC) drugs and herbals. If these are of concern in themselves or as confounders, then existing databases have limitations. There has been much discussion over the last 2 days about what might be done to improve the linkage of existing databases to increase the range of information collected. Existing databases have considerable value at present, but questions of their future additional values remain to be answered.

Phase 2 is a little more complicated and therefore probably more controversial. The goal is to identify "lesser" teratogens, those with lesser risk than thalidomide or Accutane®, and to estimate their ranges of risk or uncertainty. To do this, I would propose case-control surveillance, which is not the same as conducting individual case-control studies. In case-control surveillance, a wide range of malformations and a wide range of exposures are identified and associations between them are assessed. It is a design that has proven value and validity; we have been using case-control surveillance to study risk factors for birth defects since 1976, and a similar approach is being used by the Centers for Disease Control (CDC) in its Centers for Birth Defects Research and Prevention. Case-control surveillance has the advantage of being able to deal with the otherwise unmanageable issues related to sample size. The design does raise the concern of recall bias but, from experience in doing these studies over the years, we have learned how to enhance recall and obtain good data by asking the questions carefully. In this design, mothers of infants with various birth defects and mothers of infants without birth defects are interviewed about their exposures during pregnancy. This would be an ongoing process with no set number of interviews to be done. The interviews would accumulate over time, broadening the scope of the study. I think the CDC Centers provide an infrastructure for this that has great potential. The approach, again, could yield one of three results concerning lesser teratogenic effects: (1) drug A is definitely associated with outcome Z; (2) there is no evidence of association of drug A with outcome Z; or (3) there might be some association of drug A with outcome Z, but it is unclear. In all three situations, the range of risk that is compatible with the available data is estimated. If there is no evidence of association with an outcome initially, monitoring can continue as data accumulate to establish the range of uncertainty around that evidence. If it is unclear whether there is an association, a focused study to examine the issue more closely could be considered.

It is important to recognize that the case-control surveillance design is by no means a complete surveillance system. It is an approach to monitoring major structural malformations. It does not ascertain other outcomes such as mental retardation, or those such as autism that do not manifest for years, beyond the time when a mother could reliably recall what medications she took during pregnancy several years earlier. It is also relatively insensitive to constellations of minor defects, and thus will not identify clinical syndromes, though careful clinical involvement in the description and classification of affected children offers some opportunity to identify such constellations. Another limitation is that lesser risks of uncommonly used drugs will take a long time to identify and could entirely escape detection. That is an inescapable reality.

So, the two-step approach to surveillance for teratogenic drugs, as outlined, will detect major teratogens quickly and will detect lesser teratogens systematically and efficiently. It takes advantage of the strengths of different study designs, and its power of detection reflects the prevalence of use of the drugs and the magnitude of their risks. The strategy is compatible with regulatory, public health, medical, and patient interests.

Here are some examples of how this might work using our current knowledge of specific drugs with different levels of risk as a guide:

1. Thalidomide: Had we had this system 40 years ago when thalidomide was first introduced, when we knew nothing about its effects in humans, we would have begun by following a cohort of no more than 100 women. And we would have found very quickly that it is a major teratogen. That information could then have guided regulatory actions, patient counseling, and other activities.

2. Valproic Acid: If we accept that valproic acid increases the risk of neural tube defects by four- or

fivefold, a lesser risk than thalidomide but still one we would want to identify, then a cohort of 100 exposed pregnancies might yield 2 neural tube defects. This is the kind of result that is worrisome but not conclusive. It would be appropriate to conduct another cohort of 100. But at the same time, we could also evaluate valproic acid through case-control surveillance, which would be able to identify the four- or fivefold increased risk of neural tube defects with valproic acid.

3. Pseudoephedrine: This is a common decongestant that is widely used in pregnancy because of its OTC label and its inclusion in cough and cold preparations. A cohort of 100 pregnancies exposed to pseudoephedrine would show no evidence of increased risk for major malformations. If we then examined risks in the case-control format, we might obtain a questionable result. So what would we do? We would continue to monitor it as the case-control surveillance data accumulate, but we might also consider conducting a focused study. This is what was actually done with pseudoephedrine and one specific defect. The point is that an adverse effect might escape detection in the first phase, might be suggested in the second phase, and then could be followed in greater detail. But we would have to have a high index of suspicion and focused attention.

4. Acetaminophen: I think most of us would agree that acetaminophen is probably not a teratogen. If it were, we would have had a huge epidemic of some birth defect by now. If we were to monitor acetaminophen using this system, no adverse effect would be observed in the initial cohort. And, if we used case-control surveillance to look at the relationship with a variety of defects—cleft lip and cleft palate, tracheoesophageal fistula, and gastroschisis, for example—we still would not see evidence of risk. So then what do we do? We establish the range of risks for each of these outcomes and continue to monitor as the data accumulate. For example, if we looked at tracheoesophageal fistula in relation to acetaminophen, we might see an odds ratio of 1.0 with a confidence interval of 0.8 to 8.0. That states that the risk that is compatible with the available data is somewhere between a 20% protective effect and an eightfold increase. That is our range of uncertainty, or range of certainty, surrounding our best estimate that there is no increased risk. As more data are collected, and assuming that the risk does not change, these confidence bounds will get narrower and ultimately we might be able to say that the risk of acetaminophen is compatible with no more than a twofold increase in tracheoesophageal fistula. (Larger numbers could result in tighter confidence intervals, but concerns about bias and confounding might limit our interpretation of the validity of such tight bounds.) That is a valuable statement, and the system has the capacity to reach that conclusion.

5. Lithium: While there is not uniform agreement, we will accept for the moment that lithium increases the risk of Ebstein's anomaly, a cardiac defect, tenfold from a baseline risk of 1 in 10,000 to 1 in 1,000. To put this in perspective, taking lithium during pregnancy would increase a woman's risk of having a child with any major malformation from 3% to 3.001%. But if we evaluated lithium using this system, there would be no indication in phase 1 that it was teratogenic, and in phase 2 it is possible that we might observe a risk estimate for Ebstein's of 1.0 with a confidence interval of 0.7 to 13. That does not necessarily mean that lithium is safe, even though the observed risk is 1.0. It only means that the best we can determine is that the risk lies somewhere within the confidence interval around that point estimate with a maximum bound of 13, which happens to include the "true" risk tenfold. And so, a woman could be told that her risk of taking lithium during pregnancy is associated at most with a thirteenfold risk of this rare cardiac defect, and that is informative.

In putting all of this together, I think we need to maintain focus on what we are trying to accomplish with the information we generate. It would be possible to develop incredibly complex data sets and to create a variety of new systems. But I would argue that the basic systems that are needed already exist.

Thoughts on an Organized Approach to Post-Marketing Surveillance for the Future

Hugh Tilson, MD, DrPH
University of North Carolina School of Public Health

An Imaginary Report from the Future (tongue-in-cheek)

It is the year 2010. Underway in Bethesda, Maryland, is the 10th annual meeting of the National Joint Council on Safer Therapeutics and Medicines in Pregnancy (STAMP), which received the stamp of approval from Congress in 2005 with the U.S. Food and Drug Administration Modernization Act, Part 2. The establishment of the National Joint Council has been an extraordinary testimony to the strong national advocacy and grassroots efforts to do something about the epidemic of birth defects. The council is managed by the Centers for Education and Research in Therapeutics (CERTs) based on their existing infrastructure under the mandate of the Agency for Health Care Research and Quality (still known as AHRQ because of its expertise at two-by-two tables). It is coordinated with the Centers for Disease Control and Prevention (CDC), to which Congress appropriated $400 million for cooperative activities with AHRQ and the U.S. Food and Drug Administration (FDA). The National Joint Council's motto is "STAMP out preventable adverse pregnancy outcomes."

As always, this 10th meeting has been an outstanding and busy conference, with some very "tenth" moments. The meeting participants represent a "Who's Who" in the birth defects prevention and healthy pregnancy movement from federal agencies, academia, industry, and strong representation from consumers and patients. Two new commissioners have been named, one from the gene-splicing sector and the other a mediator to replace the lawyer in conjunction with the Tort Reform Act of 2004.

A pivotal public agenda at this meeting has been to revisit and update the Declaration of Bethesda from the National Joint Council's first meeting in the year 2000. The deliberations have reaffirmed and strengthened the five guiding principles of the council, known as "The 5 Ts":

1. Utility—whatever we do, it has to be useful, particularly to doctors and patients.
2. Priority—whatever we do, we must recognize that we cannot do it all.
3. Dignity—whatever we do, we have to remember what it is all about.
4. Equity—whatever we do, we should not single out our favorite stigmatized drug, drug company, sector, patient group, or problem.
5. Accountability—whatever we do, we must be committed to continuous process improvement.

Much of the meeting has involved focused work group sessions of the Health Action Therapeutic Subcommittees (so called because of the many HATS each must wear). The Infrastructure Work Group has accomplished the goal of establishing, by the year 2010, full national coverage of locally responsive but nationally networked Centers of Excellence, based on the network of the CERTs program. There are now 28 programs based in state health departments and 5 regional programs for those that wish to collaborate across state boundaries. These have 250 local partners in major medical centers affiliated through their local public health departments. All five regional centers have now been accredited as Pregnancy Problem Prevention Programs by the National Joint Council, and 20 state-based centers have gone through this rigorous science-based accreditation program as well. Florida was added this year after the second recount of the hard copies of the polling ballots. All the centers are now regionalized and recognized by the FDA as part of that agency's devolved, restructured, revitalized State-based Surveillance Centers of Excellence for Drug Safety (SCEDs). This revolutionary approach was based on the call for such state-based systems in the landmark Institute of Medicine (IOM) report "To Err is Human," and the subsequent evidence-based advocacy of the IOM's clinical research roundtable for the "right end" of the clinical research spectrum—community based prevention research. This couples the experience of states in conducting surveillance with the FDA's adverse drug reaction and patient safety monitoring. It has been a remarkable decade of

progress in establishing the needed infrastructure, and right on SCED-ule at that.

There is also exciting news on the automation front. One of the commitments of the National Joint Council was to harness what is known about systems and remove the busy doctor and patient from the reporting loop. The Systemwide Mandated Ultrasound Reporting Function (SMURF), which allows computerized prenatal ultrasound results to be reported directly into the surveillance system, is now in place. It is no longer necessary to ask someone to remember to report when there is an abnormal ultrasound. The Database Infrastructure Work Group has made remarkable progress. It was predicted at the year 2000 meeting that the Clinicians' Handheld Interactive Comprehensive Knowledge Exchange Network (CHICKEN) would soon be online to give doctors immediate access to their own databases and immediate input into the surveillance database system while they are seeing patients. Patients now expect that. The data are immediately routed to the Epidemiology Group for Gestational Studies (EGGS) so it is possible to know when there is a problem in a community as it develops. In other words, we know that the chicken did precede the eggs in this regard.

One objective of the nation's official health plan, "Healthy People 2010," was that information on more than 20 million covered lives in the United States be available in Large Automated Multipurpose Population-based Systems (LAMPS) to permit and facilitate automated approaches to patient-based drug safety and other surveillance. As of this meeting, there are actually 40 million covered lives available for patient-based surveillance, including the early detection of pregnancy exposures and birth defects. The World Health Organization has achieved its parallel commitment, put forth at its 20th anniversary meeting in 1999, that 20 million covered lives in Europe would be available for similar automated surveillance by the year 2010. This is the result of revolutionary work performed in Sweden, where the rich resources of nationwide pharmacy, clinical, and hospital databases are finally linked and able to "talk with each other," And, with the adoption of the database good practices standards pioneered by the International Society for PharmacoEpidemiology (ISPE), there are now standards for database work. Solid progress is in evidence on the commitment to balance the public's urgent need for such data with the mandates to ensure patient privacy.

Progress by the Workforce Work Group has not been as great as anticipated, but it is coming along. There continues to be a relative lack of public funding for training in this urgently needed field of perinatal pharmacoepidemiology, but pharmaceutical industry funding of fellowships has remained strong. The National Institutes of Health has now recognized the need for epidemiology staff development and training grants and has incorporated research, training, and career development in epidemiology as a fundamental science in its general clinical research centers program.

The Health Professions Curriculum Reform Work Group has made good progress. Epidemiology is now properly taught in 40% of health sciences schools, up from 20% in 2000, and one medical school was actually put on probation for failing to adequately teach preventive medicine. The American Academy of Pharmaceutical Physicians' Academic Fellowship program to properly train physicians in pharmaceutical medicine is in place in the expanded CERTs network, and there are now over 200 physicians who are board-certified in pharmaceutical medicine. Similar progress in schools of public health and schools of pharmacy was reported, with three new interinstitutional programs based on the University of North Carolina model reported this year.

The Public-Private Partnership Work Group has made it clear that the role of public funding is to provide core support for the infrastructure. However, industry remains strongly committed to providing private funding to ensure, through its stewardship role, that drug safety surveillance and research are properly conducted to industry standards and with industry's full partnership.

The Ethics Work Group is currently developing an academic bias and ego quantification scale and, as always, is extremely cautious about patient safety and patient dignity in the institutional review board process. The Research Agenda Work Group reports progress on risk comprehension and the communication of functional uncertainty, but these are still

problematic. That is probably good, however, because epidemiologists like to study them. Risk management strategies and continuous quality improvement methods are going well. The database data miners have undergone a "sea change" from the old Bayes approach. That is, they are using properly validated mainstream surveillance epidemiologic and accountability methods to obtain information from quality assured databases.

With all of this, however, the National Joint Council is not only about form and forum. It is the function that matters most. In her State of the Neonatal Nation Address, the council president reaffirmed the basics: It is not possible to eliminate all uncertainty and all risk, and no one wants to eliminate new drugs. She celebrated 50 new chemical entities, including one each for emesis gravidarum, gestational diabetes, and irritable bowel syndrome; a fast-acting antidepressant; a third generation HIV vaccine; and, finally, an Alzheimer's preventive medication. All of these will be used widely by women of childbearing age and thus qualify for the Prioritized Intensive Database (PID) Surveillance System. Six have been referred to intensive case-control follow-up and another six to the registry group now known as the Systematic Therapeutic Area Registry Study (STARS), another stellar example of the progress in the field possible only through the far-sighted vision of that group assembled in the year 2000.

And, finally, there has been success in the Editors' Deportment Improvement Training (EDIT) program of the National Joint Council. Editors now openly recognize that negative findings in a study represent positive findings of the absence of an adverse outcome, and not negative findings. The council's agenda that at least 50% of such articles will be accepted for publication has now been 50% accomplished. That is, 25% of these articles were accepted last year by the Journal of Unexciting Negative New Knowledge (JUNNK). The agenda of the Committee On Meta-analysis for Epidemiology is COME-ing along nicely, with growing insight into the importance of studies that do NOT find separation of end points.

Ladies and gentlemen, in short, this over-long but long-overdue update should be most up-lifting for those of us who have done the heavy lifting over the past decade.

Council participants, in your multiple responsibilities to protect the public, the mother, the baby—indeed, ALL of us—you must wear many hats yourselves. Therefore, as ever, MY hat is off to you. I will look forward to visiting you at the twentieth anniversary meeting and reporting, once again, on the further exciting progress toward our vital shared public health goals.

List of Participants

Syed Rizwanuddin Ahmad, MD, MPH
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

Elizabeth B. Andrews, PhD, MPH
Worldwide Epidemiology
GlaxoSmithKline
Research Triangle Park, NC

Marietta Anthony, PhD
Department of Pharmacology
Georgetown University Medical Center
Washington, DC

Wanda D. Barfield, MD, MPH
Division of Reproductive Health
Centers for Disease Control and Prevention
Atlanta, GA

Lisa Begg, PhD, RN
Office of Research on Women's Health
National Institutes of Health
Bethesda, MD

Ulf Bergman, MD, PhD
Huddinge University Hospital
Karolinska Institutet
Stockholm, Sweden

Mary C. Blehar, PhD
Chief, Women's Mental Health Program
National Institute of Mental Health
Bethesda, MD

Coleen A. Boyle, PhD
National Center on Birth Defects and
Developmental Disabilities
Centers for Disease Control and Prevention
Atlanta, GA

Carol Bruce, MPH, BSN
Division of Reproductive Health
Centers for Disease Control and Prevention
Atlanta, GA

Patricia Byrns, MD
Office of Research and Faculty Affairs
University of North Carolina
Chapel Hill, NC

Eileen Carey, PhD
Dupont Pharmaceuticals Company
Wilmington, DE

Christina Chambers, MPH
California Teratogen Information Services
University of California San Diego
San Diego, CA

Elizabeth Ann Conover, MS, APRN
Nebraska Teratogen Information Service
University of Nebraska Medical Center
Omaha, NE

Lois A. Corbin, RN, BSN
Safety Surveillance Department
Eli Lilly and Company
Indianapolis, IN

José F. Cordero, MD, MPH
Director, National Center on Birth Defects and
Developmental Disabilities
Centers for Disease Control and Prevention
Atlanta, GA

Adolfo Correa, MD, PhD
National Center on Birth Defects and
Developmental Disabilities
Centers for Disease Control and Prevention
Atlanta, GA

Janet Cragan, MD
National Center on Birth Defects and
Developmental Disabilities
Centers for Disease Control and Prevention
Atlanta, GA

Alan Dean, MB, BS, FFPM
Regeneration House
Kings Cross
London, England

Virgil C. Diaz, PharmD
Boehringer Ingelheim Pharmaceuticals
Ridgefield, CT

Charlotte Dickinson, PhD
National Center on Birth Defects and
Developmental Disabilities
Centers for Disease Control and Prevention
Atlanta, GA

Peggy A. Doi
PharmaResearch Corporation
Wilmington, NC

Mei-Sheng Duh, ScD
Epidemiology Division
Ingenix Pharmaceutical Services
Newton Lower Falls, MA

Dave Erickson, DDS, PhD
National Center on Birth Defects and
Developmental Disabilities
Centers for Disease Control and Prevention
Atlanta, GA

Loretta Finnegan, MD
Office of Research on Women's Health
National Institutes of Health
Bethesda, MD

Jan M. Friedman, MD, PhD
Department of Medical Genetics
University of British Columbia
Vancouver, British Columbia, Canada

Karin Galil, MD, MPH
National Immunization Program
Centers for Disease Control and Prevention
Atlanta, GA

Kathy Gallagher, DSc, MPH
Division of HIV/AIDS Prevention–Surveillance
and Epidemiology
Centers for Disease Control and Prevention
Atlanta, GA

Yvonne T. Green, RN, CNM, MSN
Director, Office of Women's Health
Centers for Disease Control and Prevention
Atlanta, GA

Nancy S. Green, MD
Medical Director
March of Dimes
White Plains, NY

Michael F. Greene, MD
Director of Maternal Fetal Medicine
Massachusetts General Hospital
Boston, MA

William W. Gregory, PhD
Pfizer Inc.
New York, NY

Marion F. Gruber, PhD
Center for Biologics Evaluation Research
Food and Drug Administration
Rockville, MD

Holli Hamilton, MD, MPH
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

James W. Hanson, MD
Division of Cancer Control and Population
Sciences
National Cancer Institute
Bethesda, MD

Janet R. Hardy, MPH, MSc
Doctoral Candidate in Epidemiology
Yale University
St. Louis, Missouri

Lewis B. Holmes, MD
Genetics and Teratology Unit
Massachusetts General Hospital
Boston, MA

Margaret Honein, PhD, MPH
National Center on Birth Defects and
Developmental Disabilities
Centers for Disease Control and Prevention
Atlanta, GA

Heddy Hubbard, RN, MPH
Center for Outcomes and Effectiveness Research
Agency for Healthcare Research and Quality
Rockville, MD

Ken Lyons Jones, MD
Division of Dysmorphology and Teratology, Department of Pediatrics, University of California San Diego Medical Center
San Diego, CA

Judith K. Jones, MD, PhD
The Degge Group, Ltd
Arlington, VA

Wanda K. Jones, DrPH
Deputy Assistant Secretary for Health Office on Women's Health
Department of Health & Human Services
Washington, DC

John M. Karon, PhD
Division of HIV/AIDS Prevention–Surveillance and Epidemilogy
Centers for Disease Control and Prevention
Atlanta, GA

Robert J. Kavlock, MD
President, Teratology Society
Reston, VA

Dianne L. Kennedy, RPh, MPH
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

Carol A. Kimmel, PhD
National Center for Environmental Assessment
Environmental Protection Agency
Washington, DC

Gary L. Kimmel, PhD
Environmental Protection Agency
Washington, DC

Susan King, MD
Visiting Scientist
National Immunization Program
Centers for Disease Control and Prevention
Atlanta, GA

Winslow Klaskala, PhD
Medical Science Principal
Hoffmann La Roche
Nutley, NJ

Sandra Kweder, MD
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

Cynthia LaCivita, PharmD
American Society of Health-System Pharmacists
Bethesda, MD

Zita Lazzarini, JD, MPH
University of Connecticut School of Medicine
Farmington, CT

Ellice Lieberman, MD, DrPH
Bringham and Women's Hospital
Harvard Medical School
Boston, MA

Mary Lou Lindegren, MD
National Center for Environmental Health
Centers for Disease Control and Prevention
Atlanta, GA

Kathleen MacQueen, PhD
National Center for HIV, STD, and TB Prevention
Centers for Disease Control and Prevention
Atlanta, GA

Russell Mardon, PhD
Supervisor of Environmental Health Surveillance
Florida Department of Health
Tallahassee, FL

Karine Martirosyan, MD, PhD
Intern, Office on Women's Health
Department of Health and Human Services
Washington, DC

Carolyn McCloskey, MD, MPH
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

John McLane, MD
Medical Science and Safety for Accutane®
Roche Labs, Inc.
Nutley, NJ

Margaret Miller, PhD, DABT
Office of the Commissioner
Food and Drug Administration
Rockville, MD

James L. Mills, MD
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, MD

Allen Mitchell, MD
Slone Epidemiology Division
Boston University Schools of Public Health and Medicine
Brookline, MA

Robert Mittendorf, MD, DrPH
Department of Obstetrics and Gynecology
Loyola University Medical Center
Maywood, IL

Godfrey Oakley, MD
Rollins School of Public Health
Emory University
Atlanta, GA

Richard Platt, MD, MS
Harvard Pilgrim Health Care
Harvard Medical School
Boston, MA

Dee Quinn, MS
Arizona Teratology Information Program
University of Arizona
Tucson, AZ

Wayne A. Ray, PhD
Department of Preventive Medicine
Vanderbilt University School of Medicine
Nashville, TN

Donna Rickert, MA, DrPH
Director of Maternal and Child Health Epidemiology
South Carolina Department of Health and Environmental Control
Columbia, SC

Aylin Riedel, PhD
Ingenix Pharmaceutical Services
Eden Prairie, MN

Evelyn M. Rodriguez, MD, MPH
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

Lewis Roht, MD
Aventis Pharmaceuticals
Kansas City, MO

Charles Schwamlein, MD, MPH
Director, Epidemiology/Medical Services
Abbott Laboratories
Abbott Park, IL

Tom Shepard, MD
Department of Pediatrics
University of Washington
Seattle, WA

Kristine Shields, MSN, MPH
Merck Pregnancy Registry Program
Merck & Company, Inc.
West Point, PA

Csaba Siffel, MD
National Center on Birth Defects and Developmental Disabilities
Centers for Disease Control and Prevention
Atlanta, GA

Catherine Sigler, DVM, PhD
Director, Epidemiologic Resources
Pfizer, Inc.
Ann Arbor, MI

Judy A. Staffa, PhD, RPh
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

Annette Stemhagen, PhD
Vice President, Strategic Development Services
Covance Periapproval Services Inc.
Radnor, PA

Lisa Summers, CNM, DrPH
Senior Technical Advisor, Professional Services
American College of Nurse-Midwives
Washington, DC

Ana Szarfman, MD, PhD
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

Sonia Tabacova, MD, PhD
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

Patricia Tennis, PhD
Worldwide Epidemiology
GlaxoSmithKline
Research Triangle Park, NC

Mary Teter, DO
Director, Drug Safety and Pharmacovigilance
Bristol-Myers Squibb & Company
Pennington, NJ

Hugh H. Tilson, MD, DrPH
Senior Advisor to the Dean
University of North Carolina
School of Public Health
Chapel Hill, NC

Anne E. Trontell, MD, MPH
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

Amarilys Vega, MD,MPH
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

Marion S. Verp, MD
Department of Obstetrics and Gynecology
University of Chicago
Lincolnwood, IL

Robert M. Ward, MD
University Medical Center
University of Utah
Salt Lake City, UT

Randolph Warren,
Thalidomide Victims Association of Canada
London, Ontario, Canada

Heather Watts, MD
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, MD

Sheila R. Weiss, PhD
Assistant Professor
University of Maryland
Baltimore, MD

Suzanne West, PhD
Cecil G. Sheps Center for Health Services Research
University of North Carolina
Chapel Hill, NC

Bruce A. Williams
Celgene Corporation
Warren, NJ

Mary Willy, PhD
Center for Drug Evaluation Research
Food and Drug Administration
Rockville, MD

Robert P. Wise, MD
Center for Biologics Evaluation Research
Food and Drug Administration
Rockville, MD

Katherine L. Wisner, MD, MS
Case Western Reserve University
Cleveland, OH

Susan F. Wood, PhD
Director, Office of Women's Health
Food and Drug Administration
Rockville, MD

Mark Yerby, MD
Oregon Health and Science University
Portland, OR

Workshop Summary

Concepts and Strategies to Actively Monitor the Risks of Medications in Pregnancy: Enhancing Post-Marketing Surveillance

November 29 - 30, 2000
Holiday Inn, Bethesda, Maryland

Department of Health and Human Services
Centers for Disease Control and Prevention
National Center on Birth Defects and Developmental Disabilities